

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problems Mailbox.**

[Title of the Invention]

**Data Analyzing Apparatus and Data Analyzing Method**

[Abstract] (amended)

[Subject] Estimation of the reliability degree and  
 5 accuracy degree are performed for data mining results,  
 and quantative estimation is performed for effective  
 factors on and after the second order.

[Means for Solving the Problems] Rules existing among  
 a plurality of data are extracted by a data mining method,  
 10 thereby outputting the rules and information about the  
 reliability degree of the rules. The S ratio and t value,  
 which are set division estimation values representing  
 the clarification degree of division when a set of a  
 plurality of data is divided into two, are defined as  
 15 the information about the reliability degree of the  
 rules.

$$S \text{ ratio} = \langle (S1+S2)/2 \rangle / S0$$

The following t values indicate if a significant  
 20 difference exists in the dispersion of the divided set  
 or if it does not exist.

$$t = \frac{|\bar{X1} - \bar{X2}|}{\sqrt{\frac{S1+S2}{N1+N2-2} \times \left( \frac{1}{N1} + \frac{1}{N2} \right)}}$$

or

$$t = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{S_1}{N_1} + \frac{S_2}{N_2}}} \quad \text{である。}$$

5

1 ( S 0 : 分割前の集合の目的変数の平方和、  
 2 S 1, S 2 : 分割後の各集合における目的変数の平方和、  
 3 N 1, N 2 : 要素数、  
 4  $\bar{X}_1, \bar{X}_2$  : 平均

- 10 1 Sum of squares of object variables of a set that  
is not divided
- 2 Sum of squares of object variables of each set that  
is divided
- 3 The number of elements
- 15 4 Average

The smaller S ratio is or the larger t value is,  
a set is precisely divided.

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-306999

(P2001-306999A)

(43) 公開日 平成13年11月2日 (2001.11.2)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	データコード (参考)
G 0 6 F 19/00	1 3 0	G 0 6 F 19/00	1 3 0
17/60	1 0 6	17/60	1 0 6

審査請求 未請求 請求項の数10 O L (全 30 頁)

(21) 出願番号 特願2000-284578 (P2000-284578)

(22) 出願日 平成12年9月20日 (2000.9.20)

(31) 優先権主張番号 特願2000-41896 (P2000-41896)

(32) 優先日 平成12年2月18日 (2000.2.18)

(33) 優先権主張国 日本 (J P)

(71) 出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番1号

(72) 発明者 津田 英隆

神奈川県川崎市高津区坂戸3丁目2番1号  
富士通エルエスアイテクノロジー株式会社  
内

(74) 代理人 100104190

弁理士 酒井 昭徳

最終頁に続く

(54) 【発明の名称】 データ解析装置およびデータ解析方法

(57) 【要約】 (修正有)

【課題】 データマイニング結果に対して、信頼度や精度の評価、2次以降の効果のある要因の定量的な評価をおこなうこと。

【解決手段】 データマイニングにより複数のデータ間\*

\*に存在するルールを抽出し、そのルールとそのルールの信頼度をあらわす情報を出力する。ルールの信頼度をあらわす情報として、複数のデータよりなる集合を2分割する際の分割の明確度をあらわす集合分割評価値S比とt値を定義する。

$S \text{ 比} = ((S_1 + S_2) / 2) / S_0$   
分割された集合の分散に有意差がない場合とある場合でそれぞれ、

$$t = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{S_1 + S_2}{N_1 + N_2 - 2} \times \left( \frac{1}{N_1} + \frac{1}{N_2} \right)}}$$

または、

$$t = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{S_1}{N_1} + \frac{S_2}{N_2}}} \quad \text{である。}$$

( $S_0$ : 分割前の集合の目的変数の平方和、  
 $S_1, S_2$ : 分割後の各集合における目的変数の平方和、  
 $N_1, N_2$ : 標本数、  
 $\bar{X}_1, \bar{X}_2$ : 平均)

S比の値が小さいほど、またt値の値が大きいほど、集合分割が明確におこなわれる。

## 【特許請求の範囲】

【請求項1】 複数のオリジナルデータよりなるオリジナルデータ群と、

前記オリジナルデータ間に存在するルールを抽出し、前記ルールの信頼度をあらわす情報を付加した信頼度情報付きルールファイルを作成するデータ処理部と、

前記ルールファイルに基づいて前記オリジナルデータの解析をおこなう解析ツールと、

を備えたことを特徴とするデータ解析装置。

【請求項2】 前記データ処理部は、前記オリジナルデータ群における各レコードの項目間の因果関係または影響度を明らかにし、隠された情報または規則性を含むルールを統計的手法により抽出することを特徴とする請求項1に記載のデータ解析装置。

【請求項3】 前記データ処理部は、前記解析ツールにおいて解析すべき項目またはその組み合わせに関する指針を含むルールをデータマイニング技法により抽出することを特徴とする請求項1に記載のデータ解析装置。

【請求項4】 前記データ処理部は、前記ルールの信頼度をあらわす情報として、複数のオリジナルデータより\*20

\*なる集合を2分割する際の分割の明確度をあらわす集合分割評価値を求めることを特徴とする請求項1に記載のデータ解析装置。

【請求項5】 前記オリジナルデータは、変動原因または変動パターンを探ることを目的とする目的変数を含み、

前記集合分割評価値は、つぎの式であらわされるS比の値であることを特徴とする請求項4に記載のデータ解析装置。

$$S \text{ 比} = ((S1 + S2) / 2) / S0$$

ただし、S0は、複数のオリジナルデータよりなる集合を2分割する前の目的変数の平方和、S1およびS2はそれぞれ分割後の各集合の目的変数の平方和である。

【請求項6】 前記オリジナルデータは、変動原因または変動パターンを探ることを目的とする目的変数を含み、

前記集合分割評価値は、つぎの式であらわされるtの値であることを特徴とする請求項4に記載のデータ解析装置。

【数1】

$$t = \frac{| \bar{X1} - \bar{X2} |}{\sqrt{\frac{S1 + S2}{N1 + N2 - 2} \times \left( \frac{1}{N1} + \frac{1}{N2} \right)}}$$

または、

$$t = \frac{| \bar{X1} - \bar{X2} |}{\sqrt{\frac{S1}{N1} + \frac{S2}{N2}}}$$

ただしS1およびS2は、それぞれ、複数のオリジナルデータよりなる集合を2分割してできた各集合の目的変数の平方和、  
N1およびN2はそれぞれ分割後の各集合の要素数、  
X1およびX2はそれぞれ分割後の各集合の平均である。

【請求項7】 前記データ処理部は、前記ルールの信頼度をあらわす情報として、2次以降の要因または条件を含む情報を求めることを特徴とする請求項1に記載のデータ解析装置。

【請求項8】 前記データ処理部での処理対象データとして、前記オリジナルデータ群から任意のデータ種または任意のデータ項目の組み合わせを抽出する前置データ処理部をさらに備えることを特徴とする請求項1〜7のいずれか一つに記載のデータ解析装置。

【請求項9】 製造工程の歩留り低下要因を解析する装置であって、

前記オリジナルデータは、変動原因または変動パターンを探ることを目的とする目的変数と、前記目的変数の変動を説明する説明変数を含み、

前記目的変数は製造歩留りであり、前記説明変数は、使用装置履歴、試験結果、設計情報および測定データを含む変数であることを特徴とする請求項1〜8のいずれか一つに記載のデータ解析装置。

【請求項10】 複数のオリジナルデータ間に存在するルールを抽出して、前記ルールにその信頼度をあらわす情報を付加した信頼度情報付きルールファイルを作成し、前記ルールとともに前記信頼度をあらわす情報を出力する工程と、

前記ルールファイルに基づいて前記オリジナルデータの解析をおこなう工程と、を含むことを特徴とするデータ解析方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、広く産業界で取り扱われるデータ間の関連を把握し、産業上優位な結果をもたらすための有意性のある結果を抽出するデータ解析装置およびデータ解析方法に関する。たとえば、半導体製造工程において、歩留りを向上させるため、製造段階で使用された装置の履歴、試験結果、設計情報および各種測定データ等に基づいて、歩留りを低下させている要因をできるだけ速やかに見つけ出す作業がおこなわれ

る。

【0002】特に、新しい製品の開発や既存の製造工程の見直しなどでは、データ解析の効率化および高信頼性を図るため、上述した種々のオリジナルデータ群からデータ中に隠れた価値ある情報や規則性を抽出することが重要である。そして、その抽出された情報や規則性を統合して学習することにより、技術者による発見が困難な知識を発見し、歩留り低下要因の発見に活用することが可能となる。これを可能とするデータ解析の手法として、特に金融や流通等の分野において活用されているデータマイニングがある。これらの産業分野は大量データを扱うため、データマイニングの適用に適している。

【0003】

【従来の技術】図25は、データマイニングを適用していない一般的なデータ解析手法を示す概念図である。一般的なデータ解析手法では、オリジナルデータ群1の各データベース2a, 2b, ...から抽出された個々のオリジナルデータは、そのまま解析ツール群3により解析される。その解析結果に基づいて、意思決定がなされる。解析ツール群3には、たとえば統計解析コンポーネント4aや図表作製コンポーネント4bなどがある。

【0004】図26は、半導体製造工程において実施されている従来のデータ解析手法を示す概念図である。従来、半導体製造工程においては、データマイニングを適用していない一般的なデータ解析手法が用いられている。オリジナルデータ群5には、設計データ、工程データ、装置データおよび試験結果等の各データベース6a, 6b, 6c, 6dが設けられている。分析対象データ7は、それら各データベース6a, 6b, 6c, 6dから抽出されたオリジナルデータにより構成される。分析対象データ7は、データ処理装置8により処理され、低歩留り要因のデータ9が得られる。

【0005】分析対象データ7は解析手順や層別に合わせて抽出される。どのデータを抽出し、それらを用いてどのように解析するかということは、個々の技術者のそれまでのノウハウ、経験および技術などに基づいて決定される。つまり、解析をおこなう技術者の裁量に任されていることになる。また、その解析結果は、相関図、トレンドグラフまたはヒストグラムなどの形式で表示される。

【0006】一般に、半導体製造工程の歩留り低下要因を究明するため、装置間差解析がおこなわれる。図27は、装置間差解析におけるロットの流れを示す概念図であり、図28は、装置間差解析におけるロットの歩留り値を、使用した装置毎に示す箱ヒゲ図である。この箱ヒゲ図は、各製造工程毎に作成される。この装置間差解析では、各ロットの工程毎の使用装置データから、各製造工程でどの製造装置を使用すると歩留りが最も影響を受けるかということが抽出される。そして、得られた箱ヒゲ図に基づいて、最も歩留りの差が顕著な工程とその装

置の同定がおこなわれる。

【0007】

【発明が解決しようとする課題】しかしながら、上述した装置間差解析では、現在、製造工程数が数百であるため、解析にあたって著しく多大な工数を要するという問題点がある。また、差異が明確に出ない場合や条件が複雑に絡み合った場合には、判断し難い場合があるという欠点がある。

【0008】また、従来の解析技術では、主に個々の技術者のそれまでのノウハウ、経験、あるいは技術などに基づいて解析が進められるため、解析の効率や信頼性の低下は避けられない。そこで、個々の技術者のノウハウや経験や技術などに依存する割合を少なくし、各解析ツールによる解析を効率的で、かつ漏れがないように実施するための知識の学習やその学習結果を活用することができ、さらにはその結果の精度を評価することが可能なデータ解析技術が望まれる。

【0009】図29は、データマイニングの一手法である判別分析のデータ解析において用いられるレコードの構成を示す模式図である。一般に、データ解析においては、変動原因や変動パターンを探ることを目的とする変数を目的変数といい、目的変数の変動を説明する変数を説明変数という。各レコード10a, 10b, ...10iは目的変数のデータ11a, 11b, ...11iと説明変数のデータ12a, 12b, ...12iに分けられる。それぞれ何を目的変数および説明変数にするか、またどのような解析をおこなうかということによって、解析の効率や信頼度が変化する。そこで、解析の信頼度、精度を評価することが必要となってくる。

【0010】半導体製造工程の歩留りに関するデータ解析をおこなう場合には、目的変数は歩留りであり、説明変数は、使用装置履歴、試験結果、設計情報および各種測定データなどである。より効率的に解析をおこない、かつその信頼性を向上させるためには、図25に示したとおり、オリジナルデータ群と既存解析ツール群との間で、目的変数と説明変数との間の関係を明らかにする処理をおこなう必要がある。

【0011】目的変数と説明変数間の関係を明らかにする手法としてデータマイニングが有効である。図30

は、データマイニングを適用した一般的なデータ解析手法を示す概念図である。データマイニングを適用したデータ解析手法では、オリジナルデータ群13の各データベース14a, 14b, ...から抽出された個々のオリジナルデータに基づいて、データ内に潜む特徴や規則性の抽出処理（データマイニング）をおこなう装置15によりルールファイル16が作成される。

【0012】そして、各データベース14a, 14b, ...から抽出された個々のオリジナルデータは、ルールファイル16に基づいて解析ツール群17により解析される。その解析結果に基づいて、意思決定がなされ

る。解析ツール群17には、たとえば統計解析コンポーネント18aや図表作製コンポーネント18bなどがある。

【0013】歩留りデータ解析においてデータマイニングを適用した場合、データマイニング結果に基づいて歩留り向上のための対策を決定したり、対策を実施すべきか否かの判定をおこなったり、対策効果の予測をおこなったりすることになる。そのためには、データマイニング結果の定量的な評価や精度が必要となる。

【0014】データマイニングの一手法である判別分析のうち、回帰木分析は特に有効である。回帰木分析の利点の一つは、結果がわかりやすいルールとして出力されることであり、それは一般的な言語やSQL言語のようなデータベース言語であらわされる。したがって、これらの結果の信頼度、精度を有効に使い、その結果により有効な意思決定をおこなったり、行動（すなわち対策等）を起こすようにすることが可能となる。

【0015】回帰木分析について簡単に説明する。回帰木分析は、図29に示すように、複数の属性を示す説明変数とそれにより影響を受ける目的変数からなるレコードの集合を対象とし、その目的変数に最も影響を与える属性と属性値を判別するものである。回帰木分析エンジンからはデータの特徴や規則性を示すルールが出力される。

【0016】回帰木分析の処理は、各属性の属性値に基づいて集合の2分割を繰り返していくことで実現される。その集合分割の際、分割前の目的変数の平方和をS0、分割後の2つの集合の目的変数の平方和をそれぞれS1およびS2とすると、つぎの(1)式であらわされるΔSが最大となるように分割するレコードの属性とその属性値が求められる。

【0017】

$$\Delta S = S0 - (S1 + S2) \quad \dots (1)$$

【0018】ここで得られる属性と属性値は、回帰木の分岐点に対応している。以降、分割された集合についても同様な処理を繰り返し、説明変数の目的変数に対する\*

$$(\text{誤判別率}) = (\text{誤判別されたデータ数}) / (\text{全データ数}) \quad \dots (2)$$

【0024】図31は、ルールの精度評価について説明するための説明図である。図31には、Δ印のA群のデータと○印のB群のデータをX1とX2の2つの説明変数で判別した様子が示されている。図31(a)に示す線型判別関数による判別では、B群のデータのうち中央付近にある○印（符号19で示す）がA群のデータであると誤判別されるので、誤判別率は12分の1（1/12）となる。

【0025】一方、図31(b)に示すマハラノビスの距離による判別では、すべて正しく判別されるので、誤判別率はゼロとなる。データ種類や解析内容によっては、データマイニング結果についても同様な手法で評価可能であるが、データ量が多いため、サンプリングによ

\*影響を調べる。したがって、分岐の上のものほど目的変数に対する影響が大きいことになる。そして、分割された集合の目的変数の標準偏差が、あらかじめ指定された値よりも小さくなると、回帰木の分岐は停止する。このように、ΔSの値により集合分割を繰り返すことにより得られる回帰木図に基づいて、各属性における属性値による目的変数への影響が判別される。

【0019】しかしながら、解析対象が半導体のように製造プロセスに関するものでは、実際の低歩留り要因は一つに限らないことがある。また、データ解析上最も有意差ありと判定された要因が実際にはそうでなかったりする場合がある。これは、通常の回帰木分析エンジンでは、集合分割をおこなう各段階で最も有意とされたものの、すなわちΔSが最大のもののみ出力されるからである。つまり、解析の精度や信頼性が十分に得られていないわけではない。

【0020】一方、金融や流通などの分野ではデータマイニングが多く適用されている。そのような分野で扱われるデータは膨大なレコード数（たとえば、POSデータ）であるが、その説明変数の数は比較的少ないものが多い。それに対して、製造プロセスから得られるデータでは、レコード数が少ないにもかかわらず、その説明変数の数が多い。このため、データマイニングの解析結果に対して、その精度や信頼度に関して、従来のものとは根本的に異なった手法が必要である。

【0021】製造プロセスのデータに対してデータマイニングを適用する場合には、その結果により製造条件を変更したりするので、データマイニング結果の精度評価は重要となる。特に、ロット数の少ないシステムLSI製品の製造プロセスにおいては、このような信頼度や精度、あるいは2次以降の可能性のある要因に関する情報は重要である。

【0022】データマイニングに限らず、一般に多変量解析における判別分析結果として得られるルールの精度評価は、つぎの(2)式にしたがっておこなわれる。

【0023】

る方法や交差検証法(cross validation)が採用される場合もある。

【0026】これらの評価方法は、個々のデータの属性が完全に既知であり、判別分析結果が正しく判別されているか否かの判定が可能であることを前提としている。しかし、プロセスデータの性質からして真の状態を判別するのは不可能であり、誤判別率による精度評価は意味をなさない。

【0027】また、仮に何らかの前提条件を付加して誤判別率を定義したとしても、プロセスデータにはデータ数が少なく、かつ説明変数の数が多いという特殊性があるため、プロセスデータに対して従来の評価手法を適用するのは不可能な場合が多い。つまり、プロセスデータ

等のデータマイニング結果の信頼度や精度に関する知見を得る手法はいまだ確立されていない。

【0028】そこで、プロセスデータ等のデータマイニング結果に対して適用可能な評価手法を新たに設ける必要がある。さらに、得られたデータマイニング結果が必ずしも真の状態を示しているとは限らないため、2次以降の要因として他の可能性のある要因についても知見を得るのが望ましい。また、データマイニングのアルゴリズムの多くは多変量解析であり、その結果を統計的手法に不慣れなエンドユーザに納得させるのは困難であるため、わかりやすい判断基準が必要である。

【0029】さらには、プロセスデータの解析では、上述したように、使用装置履歴、試験結果、設計情報および各種測定データなどを説明変数とするが、使用装置履歴、試験結果、設計情報および各種測定データ等は異なるタイプの項目群（カテゴリ）に属するデータ種、すなわち異なる質のデータ種である。一般に、従来のプロセスデータの解析手法では、別の項目群に属する説明変数の影響を分離または小さくすることが困難であるため、単一の項目群に属する説明変数についてのみ、目的変数である歩留り値との関連を解析している。したがって、得られる各解析結果は単一の項目群に属する説明変数についてのものである。

【0030】たとえば、従来の解析手法では、説明変数を各工程の使用装置履歴とした場合、各工程の装置間差による歩留りへの影響がわかるだけである。また、説明変数を電気的特性データとした場合は、どの電気的特性データがどのような値にあるとき、歩留りにどのような影響をおよぼすかという情報が得られるだけである。つまり、従来の解析手法では、異なる項目群の説明変数にまたがったり、異なる項目群に属する説明変数間の関連性についての情報は得られない。また、ある項目群に対する解析結果と別の項目群に対する解析結果との間の関連性や、異なる項目群に属する説明変数の、目的変数に対する影響度の違いなどの情報は得られない。

【0031】したがって、従来の解析手法により、たとえば装置間差が生じていたことが判明したとしても、実際にはそれよりも電気的特性データの歩留りへの影響度の方が大きい場合や、その装置間差もノイズにより得られた解析結果ほど実際には生じていない場合などがある。このような場合、使用装置履歴に対する解析結果を参考にして、装置間差の対策をおこなったとしてもさほど有効な対策とはならない。

【0032】特に、近年、デバイス等の微細化が進み、また世の中のデータの複雑化が進んだことにより、設計や製造要因により生じる不良要因はより複雑に絡み合っており、それを反映したたとえば使用装置履歴や試験結果や設計情報や各種測定データ等も各々独立ではなく互いに複雑に絡み合っている。したがって、単純に項目群毎に歩留り値におよぼす影響を解析しても、必ずしも実

際の現象を的確にあらわしているとは限らない。

【0033】そこで、異なる項目群のデータにまたがって解析をおこない、その結果としてどの項目群に属するどの説明変数が目的変数に効いているかということをノイズの少ない状態で把握することが可能なデータ解析方法やデータ解析装置が必要となる。しかし、今のところ異なる項目群に属する説明変数間に存在する関連性や絡み具合を把握する手法はない。

【0034】この発明は、上述した事情に鑑みてなされたものであり、データマイニング結果、特に半導体製造工程において取得されるプロセスデータ等のデータマイニング結果に対して、信頼度や精度の評価、あるいは2次以降の効果のある要因の定量的な評価をおこなうことができるデータ解析装置およびデータ解析方法を提供することを目的とする。

【0035】また、本発明の他の目的は、属する項目群が異なる複数の項目を一括して並列に説明変数として取り扱い、抽出されたルールに基づく回帰木図の、注目するノードに属する集合または属さない集合に対して解析をおこなうことによって、ノイズがより少ない状態で解析をおこなうことができるデータ解析装置およびデータ解析方法を提供することである。

【0036】

【課題を解決するための手段】上記目的を達成するため、この発明は、データ処理部により、オリジナルデータ群の複数のオリジナルデータ間に存在するルールを抽出し、そのルールの信頼度をあらわす情報を付加した信頼度情報付きルールファイルを作成する。抽出されたルールは、その信頼度をあらわす情報とともに出力される。そして、解析ツールにより、前記ルールファイルに基づいてオリジナルデータの解析をおこなう。

【0037】この発明において、オリジナルデータ間に存在するルールを抽出する際に、統計的手法、たとえばデータマイニング技法を用いる。また、前記ルールの信頼度をあらわす情報として、複数のオリジナルデータよりなる集合を2分割する際の分割の明確度をあらわす集合分割評価値を求める。前記ルールの信頼度をあらわす情報は、2次以降の要因または条件を含む情報であってもよい。

【0038】また、オリジナルデータは、変動原因または変動パターンを探ることを目的とする目的変数と、前記目的変数の変動を説明する説明変数を含む。この発明を製造工程の歩留り低下要因を解析する装置に適用する場合には、目的変数は製造歩留りである。また、説明変数は使用装置履歴、試験結果、設計情報および測定データなどである。

【0039】また、集合分割評価値の一つとしてS比を設定する。複数のオリジナルデータよりなる集合を2分割する前の目的変数の平方和をS0、分割後の各集合の目的変数の平方和をそれぞれS1およびS2とすると、

10

20

30

40

50



S比はつぎの(3)式であらわされる。

【0040】

S比 =  $(S1 + S2) / 2 / S0 \dots (3)$

【0041】S比は、集合分割による平方和の低減率であり、集合分割により平方和がどの程度低減したかを示すパラメータである。この値が小さいほど集合分割の効果は大きい。すなわち、集合分割が明確におこなわれていることになるので、装置間差が大である。

【0042】また、集合分割評価値の一つとしてt値を設定する。この値は、分割後の各集合の平均の差の検定 \* 10

$$t = \frac{|\bar{X1} - \bar{X2}|}{\sqrt{\frac{S1 + S2}{N1 + N2 - 2} \times \left(\frac{1}{N1} + \frac{1}{N2}\right)}} \dots (4)$$

$$t = \frac{|\bar{X1} - \bar{X2}|}{\sqrt{\frac{S1}{N1} + \frac{S2}{N2}}} \dots (4)'$$

ただし、(4)式は分割された集合の分散に有意差がない場合に適用され、一方、(4)'式は分割された集合の分散に有意差がある場合に適用される。

【0044】t値は、2分割された集合における2つの母平均の差の検定のための値であり、分割された集合における目的変数の平均値の有意差を示す基準となる。自由度、すなわちデータ数が同じであるならば、t値が大きいほど集合が明確に分割されていることになるので、装置間差が大となる。

【0045】また、オリジナルデータの各レコードが、説明変数としてタイプの異なる複数の項目群に属する複数の項目を有しており、抽出されたルールに基づく回帰木図の、注目するノードに属する集合または属さない集合に対して解析をおこなう構成としてもよい。

【0046】この発明によれば、オリジナルデータ群の複数のオリジナルデータ間に存在するルールに、その信頼度をあらわす情報が付加された信頼度情報付きルールファイルが作成される。その作成されたルールと前記信頼度をあらわす情報はともに出力される。そして、作成されたルールファイルに基づいてオリジナルデータの解析がおこなわれる。したがって、抽出されたルールの精度評価をおこなうことができるとともに、2次以降の効果のある要因をその定量的な評価値(S比やt値)とともに得ることができる。

【0047】また、この発明によれば、異なる項目群に属する複数の項目を一括して説明変数として取り扱い、回帰木分析を実施することにより、異なる項目群にまたがって目的変数に影響をおよぼす説明変数を信頼度情報付きで得ることができる。その結果、説明変数としてどの項目群のどのデータが目的変数に対してどのように影響を与えているかという情報を得ることができる。

【0048】また、回帰木図の上位に位置する説明変数

\*に用いられる。複数のオリジナルデータよりなる集合を2分割してできた各集合の目的変数の平方和をそれぞれS1およびS2、分割後の各集合の要素数をそれぞれN1およびN2、分割後の各集合の平均をそれぞれ $\bar{X1}$ および $\bar{X2}$ (Xの前の“/”はバーを意味する)とすると、t値はつぎの(4)式および(4)'式であらわされる。

【0043】

【数2】

ほど目的変数に対する影響度が大きいと、注目するノードに属する集合または属さない集合について解析をおこなうことによって、注目するノードの項目よりも影響度が大きい項目による影響度を除いた状態で解析をおこなうことができる。

【0049】また、注目するノードよりも下位のノードに関するデータを除いたオリジナルデータについて解析をおこなうことによって、注目するノードの項目よりも影響度が小さい項目による影響度を除いた状態で解析をおこなうことができる。

【0050】

【発明の実施の形態】以下に、この発明の実施の形態について図面を参照しつつ詳細に説明する。図1は、この発明にかかるデータ解析装置の構成の一例を示す概略図である。このデータ解析装置は、それぞれ複数のオリジナルデータを含むデータベース24a、24b、...からなるオリジナルデータ群23を備える。また、データ解析装置は、オリジナルデータ群23から任意のデータ種または任意のデータ項目の組み合わせを処理対象データとして抽出する第1のデータ処理部29を備える。この第1のデータ処理部29は前置データ処理部としての機能を有する。

【0051】また、データ解析装置は、第1のデータ処理部29によりオリジナルデータ群23から抽出された個々のオリジナルデータ、またはオリジナルデータ群23に含まれるすべてのオリジナルデータに基づいて、データ内に潜む特徴や規則性などのルールの抽出処理(データマイニング)をおこなう第2のデータ処理部25を備える。第2のデータ処理部25は、ルールの抽出と

もに、そのルールにその信頼度をあらわす情報を付加した信頼度情報付きルールファイル26を作成する。

【0052】抽出されたルールは、信頼度情報とともに図示しない表示装置の画面表示や、図示しない印刷装置により印刷された帳票として出力される。そして、信頼度情報付きルールファイル26に基づいて、オリジナルデータ群23のオリジナルデータは、解析ツール群27の各解析ツール28a, 28b, ...により解析される。その解析結果に基づいて意思決定がなされる。

【0053】図2は、この発明を半導体製造工程の歩留り解析に適用した場合のオリジナルデータのデータ形式例を示す模式図である。オリジナルデータ群23から出力されるデータは、各ロットのデータを1レコードとし、各ロットの各工程における使用装置名を説明変数とする。目的変数は歩留りである。この形式のデータはデータマイニングの一手法である判別分析エンジンの入力データとなる。

【0054】以下、この発明を半導体製造工程の歩留り解析に適用し、さらにデータマイニングの判別分析として回帰木分析を適用した場合を例にして説明する。なお、回帰木分析の概要については説明済みであるため、省略する。

【0055】図3～図7は、回帰木分析プログラムの起動時または実行中の画面表示例を示す図である。図3はプログラム起動時の画面表示例、図4は分析ファイル選択時の画面表示例である。図5は、プログラム終了条件である標準偏差指定時の画面表示例である。回帰木分析の実行中に、2分割された集合の目的変数の標準偏差が、図5に示す標準偏差指定画面で指定された値よりも小さくなると、回帰木の分岐が停止する。図6は、データ処理結果であるルールを出力するフォルダ指定時の画面表示例、図7はプログラム実行中の画面表示例である。

【0056】〔基本的解析手法〕図8は、回帰木分析結果を説明するための回帰木図の一例を示す図である。図8に示す例を用いて回帰木図の見方を説明する。ただし、ここでは説明を簡略化するため、回帰木図とともに出力される信頼度情報については図示省略し、その説明を省略する。信頼度情報の説明については後述する。また、図8において、回帰木図内の各節におけるnは、該当するロット数を示す。

【0057】全ロットの平均歩留りは78.5%である。STA工程において装置MECA01を使用したロットの平均歩留りは85.7%であり、装置MECA02/03を使用したロットの平均歩留りは69.2%である。図示例では、MECA01を使用したロットの方が歩留りがよい傾向にある。ここで、MECA02/03とは、2つの装置MECA02とMECA03のいずれかを使用していることを意味する（MECC02/03も同様）。

【0058】したがって、対象とした全工程のうち、使用した装置により最も歩留りが影響を受けるのは、STA工程において3台ある装置MECA01, MECA02, MECA03のうち、MECA01とMECA02/03のいずれを使用するかということである。

【0059】そして、STA工程でMECA01を使用したロットに限れば、STC工程においてMECC01を使用したロットの平均歩留りは90%であり、MECC02/03を使用したロットの平均歩留りは80%である。したがって、MECA01を使用したロット群において、歩留りに最も大きな影響をおよぼすのは、STC工程において3台ある装置MECC01, MECC02, MECC03のうち、MECC01とMECC02/03のいずれを使用するかということである。

【0060】また、STA工程でMECA02/03を使用したロットに限れば、STB工程においてMECB03を使用したロットの平均歩留りは75%であり、MECB02を使用したロットの平均歩留りは40%である。したがって、MECA02/03を使用したロット群において、歩留りに最も大きな影響をおよぼすのは、STB工程においてMECB02とMECB03のいずれを使用するかということである。

【0061】図9は、図8に示す回帰木分析の結果をまとめて示す図表である。図9から明らかなように、図8に示す回帰木図の場合、STA工程でMECA01を使用し、かつSTC工程でMECC01を使用した場合が最も高歩留り（90%）であることがわかる。一方、STA工程でMECA02/03を使用し、かつSTB工程でMECB02を使用した場合が最も低歩留り（40%）であることがわかる。これらのことがデータマイニングの結果、すなわちルールとして判明したことになる。

【0062】得られたルールに基づいて、従来と同様にして個々の解析ツールによりオリジナルデータの解析をおこない、たとえば箱ヒゲ図（図28参照）により意思決定をおこなう。具体的には、STA工程でMECA02/03の装置を使用し、かつSTB工程でMECB02の装置を使用したロット群の歩留りと、それ以外のロット群の歩留りを比較する。また、MECA02/03, MECB02、またはそれら組み合わせ状態を調査する。このようにデータマイニング手法を適用してデータ解析をおこなうことによって、事前にその傾向、特徴、影響度等を定量的に把握することにより、より効果的にかつ信頼性の高い結果を得ることができる。

【0063】回帰木図などのルールとともに出力される信頼度情報について説明する。この信頼度情報は、評価用統計値リストとして出力される。このリストにより、たとえば回帰木分析結果の定量的な評価などがおこなえることになる。本実施の形態では、評価用統計値リストとして、特に限定しないが、全集合の目的変数（すなわ

10

20

30

40

50

ち歩留り)の平均、標準偏差およびデータ数が出力される。図10～図16に、回帰木分析結果の画面表示例を示す。

【0064】図10～図16において、画面30の下半部に評価用統計値リストの表示領域31が設けられており、その表示領域31内の「全体ave=」、「s=」および「N=」に続く各値がそれぞれ全集合の目的変数の平均値、標準偏差およびデータ数をあらわしている。画面30の上半部は回帰木図の表示領域32となっている。

【0065】さらに本実施の形態では、評価用統計値リストとして、回帰木分析の対象データを最初に2分割するときの前記(1)式の $\Delta S$ の大きさの順に、特にその数を限定しないが、たとえば10項目の評価用統計値が出力される。すなわち、前記(3)式のS比、前記

(4)式のt値、分割された集合の目的変数の平均値の差、分割された各集合のデータ数、分割された集合の属性名(ここでは工程名)、分割された2つの集合の属性値(ここでは装置名)およびその目的変数の大小関係が出力される。

【0066】図10～図16において、評価用統計値リストの表示領域31内の「S比=」、「t=」、「差」、「N」および「attr=」に続く各値がそれぞれS比、t値、分割された集合の目的変数の平均値の差、分割された各集合のデータ数、および分割された集合の属性名をあらわしている。また、評価用統計値リストの表示領域31内の「group=」に続く括弧内の表記が分割された2つの集合の属性値とその目的変数の大小関係をあらわしている。

【0067】ここで、S比は、その値が小さいほど集合分割による評価値の減少率が大きいことをあらわしており、装置間差が大となる。t値は、その値が大きいほど集合分割が明確におこなわれたことをあらわしており、装置間差が大となる。分割された集合の目的変数の平均値の差は、その値が大きいほど使用装置による歩留り差が大きいことをあらわしており、装置間差が大となる。分割された各集合のデータ数は、2つの値に近いほど異常データの影響が小さいことをあらわしており、装置間差が大となる。

【0068】回帰木作成時において評価対象となる $\Delta S$ の値は、解析対象とする各目的変数と説明変数の状態により異なる。すなわち、個々の分析エンジン実行毎に異なるので、各実行結果の精度評価をおこなうには、各実行時における結果を共通的に評価できる値が必要となる。そこで、本実施の形態では、別々に実行した結果に対して比較または評価することが可能なパラメータとしてS比およびt値が出力される。

【0069】この2つのパラメータによって、同一の実行時のみならず、別々に実行した結果に対しても比較または評価をおこなうことが可能となる。そして、複数の

回帰木図にわたって、評価用統計値リストの表示領域31内に表示された評価用統計値リストを参照し、解析することによって、より多くの知見を得ることができる。

【0070】本実施の形態では、装置間差解析において回帰木図で上の階層に表れてくる工程、または $\Delta S$ の値が大きく評価用統計値リストの上位にきている工程ほど装置間差が大である。つまり、本実施の形態では、回帰木図だけでなく評価用統計値リストに基づいて問題工程や装置を特定することができる。それに対して、従来は、集合分割の最上階層において、(1)式の $\Delta S$ に基づいて集合を分割する際に通常回帰木図に出力されるのは $\Delta S$ が最大のもののみである。

【0071】したがって、わずかな差異やノイズまたは揺らぎ等により $\Delta S$ が最大にならなかったもの、すなわち回帰木図に出力されていない工程や装置が実際の問題工程や装置である場合がある。また、不良要因が複数である場合はその識別は不可能である。

【0072】つぎに、発明者が半導体製造工程に対してこの発明を適用して解析をおこなった例を挙げて、評価用統計値リストの解釈の仕方について説明するとともに、この発明の特徴をより明らかとする。

【0073】[各品種毎の解析] まず、同一テクノロジーで同じような設計条件である5品種(MB01～MB05)に対しておこなった解析について説明する。これらの品種はほとんど同じ工程を通して処理された。

【0074】従来通り、個々の品種毎に回帰木分析を実行した結果を図10～図14にそれぞれ示す。図10～図14には、それぞれMB01～MB05の回帰木図とともに評価用統計値リストが示されている。ここで、評価用統計値リスト内の上位10項目について、上位から順に各項目毎に「S比;t値;分割された集合の目的変数の平均値の差;分割された各集合のデータ数;分割された集合の属性名;分割された2つの集合の属性値およびその目的変数の大小関係」の順に示す。なお、大小関係は「>」または「<」の記号で示す。

【0075】まず、MB01(図10参照)について、第1番目は「0.2508;6.459;30.58286;35と9;ST23;ST23M2/ST23M1>ST23M3」である。第2番目は「0.3353;4.542;20.59695;27と17;ST35;ST35M2>ST35M4/ST35M3」である。第3番目は「0.3571;4.100;22.29412;34と10;ST17;ST17M1>ST17M5」である。第4番目は「0.3811;3.619;17.49804;17と27;ST10;ST10M4>ST10M3/ST10M2」である。第5番目は「0.3837;3.569;17.78299;29と15;ST48;ST48M2/ST48M3>ST48M4」である。

【0076】また、第6番目は「0.3999;3.2

42; 37. 53333; 42と2; ST49; ST49M4/ST49M3>ST49M2」である。第7番目は「0. 4098; 3. 041; 14. 86045; 21と23; ST21; ST21M4>ST21M3」である。第8番目は「0. 4173; 2. 886; 24. 725; 40と4; ST18; ST18M2/ST18M3>ST18M1」である。第9番目は「0. 4190; 2. 849; 14. 61786; 28と16; ST20; ST20M1>ST20M2」である。第10番目は「0. 4227; 2. 770; 32. 97619; 42と2; ST44; ST44M1>ST44M2/nop」である。なお、「nop」はその工程(ここではST44)に相当するデータが欠損していたため、人為的に補充した。

【0077】つぎに、MB02 (図11参照) について、第1番目は「0. 3649; 3. 102; 13. 07326; 17と11; ST27; ST27M2/ST27M1<ST27M3」である。第2番目は「0. 3840; 2. 803; 12. 11497; 11と17; ST3; ST3M2/ST3M1>ST3M3」である。第3番目は「0. 3867; 2. 760; 12. 945; 20と8; ST30; ST30M2/ST30M1<ST30M3」である。第4番目は「0. 4042; 2. 483; 13. 10454; 22と6; ST31; ST31M1<ST31M3/ST31M2」である。第5番目は「0. 4262; 2. 121; 15. 25467; 25と3; ST39; ST39M2>ST39M1」である。

【0078】また、第6番目は「0. 4263; 2. 121; 10. 1; 9と19; ST45; ST45M2<ST45M1/ST45M3」である。第7番目は「0. 4309; 2. 042; 11. 92261; 23と5; ST43; ST43M3/ST43M1>ST43M2」である。第8番目は「0. 4341; 1. 987; 10. 3; 7と21; ST51; ST51M2<ST51M1」である。第9番目は「0. 4357; 1. 958; 14. 23867; 25と3; ST8; ST8M4/ST8M2<ST8M3」である。第10番目は「0. 4359; 1. 955; 23. 69259; 27と1; ST50; ST50M1<ST50M2」である。

【0079】つぎに、MB03 (図12参照) について、第1番目は「0. 2897; 3. 810; 15. 58333; 18と4; ST27; ST27M1/ST27M2>ST27M3」である。第2番目は「0. 2897; 3. 810; 15. 58333; 18と4; ST32; ST32M1/ST32M3>ST32M2」である。第3番目は「0. 3352; 3. 136; 10. 82137; 13と9; ST42; ST42M2/ST42M3>ST42M1」である。第4番目は「0. 3

498; 2. 931; 10. 90571; 15と7; ST48; ST48M2>ST48M3/ST48M4」である。第5番目は「0. 3596; 2. 795; 9. 863335; 10と12; ST26; ST26M4>ST26M3」である。

【0080】また、第6番目は「0. 3636; 2. 739; 9. 681816; 11と11; ST18; ST18M2>ST18M3/ST18M1」である。第7番目は「0. 3673; 2. 688; 13. 91403; 19と3; ST38; ST38M3>ST38M4」である。第8番目は「0. 3747; 2. 586; 9. 644646; 14と8; ST22; ST22M3/ST22M1<ST22M2」である。第9番目は「0. 3774; 2. 549; 13. 37368; 19と3; ST8; ST8M2>ST8M3」である。第10番目は「0. 3814; 2. 494; 9. 690475; 15と7; ST45; ST45M1/ST45M3>ST45M2」である。

【0081】つぎに、MB04 (図13参照) について、第1番目は「0. 3266; 3. 643; 35. 00326; 23と4; ST22; ST22M3/ST22M2/nop>ST22M1」である。第2番目は「0. 3380; 3. 461; 45. 892; 25と2; ST13; ST13M2/ST13M1/ST13M3>nop」である。第3番目は「0. 3380; 3. 461; 45. 892; 25と2; ST14; ST14M1>nop」である。第4番目は「0. 3380; 3. 461; 45. 892; 25と2; ST15; ST15M2>nop」である。第5番目は「0. 3380; 3. 461; 45. 892; 25と2; ST16; ST16M3/ST16M2>nop」である。

【0082】また、第6番目は「0. 3380; 3. 461; 45. 892; 25と2; ST52; ST52M1>nop」である。第7番目は「0. 3380; 3. 461; 45. 892; 25と2; ST53; ST53M1>nop」である。第8番目は「0. 3380; 3. 461; 45. 892; 25と2; ST47; ST47M2/ST47M1>ST47M3」である。第9番目は「0. 3882; 2. 683; 24. 01191; 21と6; ST31; ST31M2>ST31M1」である。第10番目は「0. 3889; 2. 672; 20. 02666; 12と15; ST46; ST46M2<ST46M3/ST46M1」である。

【0083】つぎに、MB05 (図14参照) について、第1番目は「0. 1403; 7. 161; 37. 86496; 9と13; ST3; ST3M2/ST3M3<ST3M1」である。第2番目は「0. 2584; 4. 324; 30. 64167; 10と12; ST7; ST7M4<ST7M5」である。第3番目は「0. 2878; 3. 841; 28. 59999; 11と11;

ST52; ST52M1<nop」である。第4番目は「0. 2878; 3. 841; 28. 59999; 11と11; ST53; ST53M1<nop」である。第5番目は「0. 3418; 3. 043; 24. 79833; 12と10; ST34; ST34M1<ST34M2」である。

【0084】また、第6番目は「0. 3444; 3. 006; 24. 90086; 13と9; ST46; ST46M1/ST46M2<ST46M3」である。第7番目は「0. 3522; 2. 897; 34. 77193; 3と19; ST33; ST33M3<ST33M2」である。第8番目は「0. 3555; 2. 851; 23. 99829; 13と9; ST37; ST37M2/ST37M3<ST37M4」である。第9番目は「0. 3586; 2. 808; 27. 85412; 17と5; ST28; ST28M2>ST28M1」である。第10番目は「0. 3709; 2. 638; 22. 68205; 13と9; ST29; ST29M2/ST29M3<ST29M1」である。

【0085】図10～図14に示す各回帰木図に加え、各評価用統計値リストを解釈することによって、以下のことがわかる。MB02、MB03およびMB04では、主にtの値から装置間差が小さいとみなすことができる。この結果、単独5品種の解析では、MB01の結果より、ST23M3が主な異常装置とみなすことができる。また、MB05の結果より、ST3M2/ST3M3が主な異常装置とみなすことができる。

【0086】ここで、ST23とST3の工程では同じ装置が使用されている。すなわち、ST23M3とST3M3は同じ装置である。したがって、単独5品種の解析により異常装置が同定されたことになる。また、MB04では7工程が交絡している。物理解析を含めた判断がST23M3 (ST3M3) の装置が異常であったことからすると、t値は極めて有効な評価値であることがわかる。

【0087】[全品種データの同時解析 (ST3工程を含む)] つぎに、上記5品種 (MB01～MB05) のデータをマージしておこなった解析について説明する。一般にデータマイニングには大量のデータが必要である。しかし、システムLSI製品の場合、上述したように品種毎に解析をおこなうにはデータ量が少ないことがある。そこで、異なる品種であっても同一品種のように扱って解析をおこなうことが可能であることを検証するために本解析を実施した。

【0088】MB01～MB05の品種毎のデータをマージして、回帰木分析エンジンの入力データとした。目的変数は各品種の歩留り値のままであり、説明変数は各品種の装置履歴のままであった。MB02では他の品種に比べて装置履歴が全くない工程が6個あるため、MB01およびMB03～MB05の4品種についてこれを

削除した。このようにして、異なる5品種を同一品種として扱った。図15に、回帰木分析を実行した結果得られた回帰木図および評価用統計値リストの画面表示例を示す。また、[各品種毎の解析]と同様の書式で評価用統計値リストの内容を示す。

【0089】第1番目は「0. 4260; 4. 950; 15. 83319; 70と73; ST3; ST3M1/nop>ST3M2/ST3M3」である。第2番目は「0. 4286; 4. 847; 21. 96218; 12と21; ST31; ST31M1/ST31M5<ST31M2」である。第3番目は「0. 4300; 4. 791; 15. 41398; 75と68; ST48; ST48M2>ST48M4/ST48M3」である。第4番目は「0. 4508; 3. 923; 14. 6089; 38と105; ST54; ST54M4/nop>ST54M3」である。第5番目は「0. 4528; 3. 835; 20. 63266; 128と15; ST7; ST7M4<nop/ST7M5」である。

【0090】また、第6番目は「0. 4558; 3. 696; 12. 33582; 62と81; ST33; ST33M2>ST33M3/ST33M1」である。第7番目は「0. 4666; 3. 179; 15. 67419; 124と19; ST55; ST55M2>ST55M1」である。第8番目は「0. 4697; 3. 016; 10. 13068; 73と70; ST43; ST43M1/ST43M2>ST43M3」である。第9番目は「0. 4714; 2. 927; 10. 37228; 94と49; ST17; ST17M1/nop>ST17M5/ST17M4」である。第10番目は「0. 4716; 2. 913; 12. 35196; 28と115; ST30; ST30M3/ST30M4>ST30M2」である。

【0091】評価用統計値リストにより、「ST3M1>ST3M2/ST3M3」、「ST31M2>ST31M1/ST31M5」、「ST48M2>ST48M3/ST48M4」の順でほぼ同程度の装置間差があることがわかる。これらの3工程では、S比およびt値ともほぼ同じ値であり、第4番目以降の値とは有意な差がある。特に、t値は4. 95～4. 79であり、いずれも[各品種毎の解析]で得られたMB02、MB03およびMB04の単独解析結果における最上位分岐のt値よりも大きくなっている。このt値 (4. 95～4. 79) よりも大きな分岐は、[各品種毎の解析]で得られたMB01およびMB05の単独解析結果における最上位分岐のもののみである。

【0092】したがって、本解析結果によりST23M3、すなわちST3M3が異常装置であることはより明確となり、各品種毎におこなった解析結果と同じ結果が得られたことになる。ST23M3 (ST3M3) の他にも、同程度の装置間差があることが示されたST31

M1/ST31M5およびST48M4/ST48M3の装置は【各品種毎の解析】で得られた結果においても一応候補にあげられるものである。なお、ST31の工程では、ST17およびST7の各工程と同じ装置を使用した。また、ST48の工程では、ST35の工程と同じ装置を使用した。

【0093】【全品種データの同時解析（ST3工程を削除）】つぎに、上述した【全品種データの同時解析（ST3工程を含む）】において使用したデータからST3工程を削除したデータに対しておこなった解析について説明する。図16に、回帰木分析を実行した結果得られた回帰木図および評価用統計値リストの画面表示例を示す。また、【各品種毎の解析】と同様の書式で評価用統計値リストの内容を示す。

【0094】第1番目は「0.4286;4.847;21.96218;122と21;ST31;ST31M1/ST31M5<ST31M2」である。第2番目は「0.4300;4.791;15.41398;75と68;ST48;ST48M2>ST48M4/ST48M3」である。第3番目は「0.4508;3.923;14.6089;38と105;ST54;ST54M4/nop>ST54M3」である。第4番目は「0.4528;3.835;20.63266;128と15;ST7;ST7M4<nop/ST7M5」である。第5番目は「0.4558;3.696;12.33582;62と81;ST33;ST33M2>ST33M3/ST33M1」である。

【0095】また、第6番目は「0.4666;3.179;15.67419;124と19;ST55;ST55M2>ST55M1」である。第7番目は「0.4697;3.016;10.13068;73と70;ST43;ST43M1/ST43M2>ST43M3」である。第8番目は「0.4714;2.927;10.37228;94と49;ST17;ST17M1/nop>ST17M5/ST17M4」である。第9番目は「0.4716;2.913;12.35196;28と115;ST30;ST30M3/ST30M4>ST30M2」である。第10番目は「0.4717;2.906;10.08816;89と54;ST61;ST61M1>ST61M2」である。

【0096】評価用統計値リストにより、【全品種データの同時解析】（ST3工程を含む）において最も装置間差が大きかったST3工程を除いた解析では、ST31M1/ST31M5の装置を使用したことにより歩留りが低下しているのがわかる。また、【全品種データの同時解析】（ST3工程を含む）と【全品種データの同時解析（ST3工程を削除）】を通じてST48M\*

（\*は装置の号機をあらわす数字である）を使用するST48工程は回帰木図に表れていないので、ST3工程

およびST31工程と交絡している。

【0097】また、ST3工程およびST31工程は、回帰木図と評価用統計値リストの上位に出ているので、互いに独立とみなすことができる。このように、回帰木図や評価用統計値リストを解析することにより、装置間差の独立性、交絡性の情報が得られる。ST3工程とST31工程の独立性は処理内容からみても整合性が取れている。また、ST48工程の交絡性はどちらが真の要因かを判別する調査に値する。

10 【0098】【タイプの異なる複数の項目群にまたがる解析】つぎに、オリジナルデータの各レコードが、タイプの異なる複数の項目群に属する複数の項目からなる説明変数を有する場合について説明する。図17は、回帰木分析の入力となるデータ形式の一例を示す模式図である。オリジナルデータの各レコードはウェーハ番号単位になっている。各レコードは、各製造工程（工程A、工程B）での使用装置名および各電気的特性データ（RSP、WET<sub>i</sub>）の値を説明変数とし、ウェーハの歩留りを目的変数として構成される。つまり、図17に示す例では、歩留りに効果があるのは装置間差と電気的特性データであるとする。ここで、使用装置の項目と電気的特性データの項目は異なるタイプの項目群に属する。

20 【0099】以下、図17に示すオリジナルデータに対して解析をおこなう場合を例にして説明する。図18は、図17に示すオリジナルデータに対して回帰木分析を実行した結果を示す画面表示例であり、図19は、図18に示す回帰木分析の結果をまとめて示す図表である。なお、図18中、n0～n5は、説明の便宜上つけたノード番号であり、画面に表示されるわけではない（図22も同様）。

30 【0100】図18について説明する。全ウェーハ、すなわち1000枚のウェーハの平均歩留りは74.8%である（ノードn0）。全ウェーハのうち200枚のウェーハについては、工程Aにおいて装置AM1を使用しており、その平均歩留りは54%である（ノードn1）。全ウェーハのうち残りの800枚のウェーハについては、工程Aにおいて装置AM2を使用しており、その平均歩留りは80%である（ノードn2）。ノードn1の200枚のウェーハは、工程Bにおいて、装置BM1を使用した80枚のウェーハ（平均歩留り45%）と、装置BM3を使用した120枚のウェーハ（平均歩留り60%）に2分割される。

40 【0101】ノードn2の800枚のウェーハのうち600枚のウェーハについては、工程Cにおいて装置CM2を使用しており、その平均歩留りは90%である（ノードn3）。ノードn2の800枚のウェーハのうち工程Cにおいて装置CM1を使用した200枚のウェーハの平均歩留りは50%である。ノードn3の600枚のウェーハのうち200枚のウェーハについては、電気的特性データRSPの値が90よりも大であり、その平均



歩留りは80%である(ノードn4)。一方、ノードn3の残りの400枚のウェーハについては、電気的特性データRSPの値が90以下であり、その平均歩留りは95%である(ノードn5)。

【0102】図18および図19よりつぎのことがわかる。歩留りに最も影響をおよぼすのは、工程AでAM1とAM2のうちいずれの装置を使用するかということであり、AM2の装置を使用した方が歩留りがよい。工程AでAM2の装置を使用したもののうち、工程CでCM1の装置よりもCM2の装置を使用した方が歩留りがよい。さらに、工程AでAM2の装置を使用し、かつ工程CでCM2の装置を使用したウェーハ群に対しては、電気的特性データRSPの状態が最も歩留りに影響をおよぼす。一方、工程AでAM1の装置を使用したものについて、そのつぎに歩留りに影響をおよぼすのは工程Bで使用した装置である。

【0103】このように、図18に示す例の回帰木図においては、上位2階層はいずれも装置間差によるものである。回帰木図では、上階層に現れる説明変数ほど目的変数に対する影響が大きいので、この例では全ウェーハを用いた解析では歩留りに影響が大きいのは複合条件を含めても装置間差であり、電気的特性データはあまり効いていないように見られる。しかし、工程AでAM2の装置を使用し、かつ工程CでCM2の装置を使用したウェーハ群については、歩留りに最も効くのは電気的特性データRSPである。このように、全ウェーハの平均歩留りは74.8%であるが、使用装置や電気的特性データとの関連で幾つかの集合に分けることによって、上述したような特徴や規則性が存在することがわかる。回帰木分析は、このような特徴や規則性を自動的に抽出し、歩留り解析の手がかりを与える。

【0104】図20は、図17に示すオリジナルデータに対して、電気的特性データRSPと歩留りとの相関関係を示す特性図であり、同図(a)は図18のノードn3の600枚のウェーハについての相関図であり、同図(b)は図18のノードn0の1000枚のウェーハについての相関図である。図20(a)に示す相関図では、電気的特性データRSPよりも大きな影響を歩留りにおよぼす工程Aおよび工程Cの装置間差の影響が除去されている。それに対して、図20(b)に示す相関図では、工程Aおよび工程Cの装置間差の影響が含まれているため、同図(a)に示す相関図よりもノイズが多くなっている。

【0105】したがって、図18に示す回帰木図を基にして、ノードn3の600枚のウェーハを選抜し、それらについて解析をおこなうことにより、電気的特性データRSPが歩留りに与える影響をより正確に知ることができる。このように回帰木分析は、歩留り-電気的特性データRSPの相関関係を求める際に、どの装置使用条件のデータに限って解析をおこなえばよいかということ

を自動的に抽出する。たとえば、回帰木分析の一つの出力情報として各ノードに属するレコードの識別番号(レコードid)を出力するようにしておき、その抽出された識別番号の情報を従来の解析ツールに読み込ませることにより、各装置使用条件毎に歩留り-電気的特性データRSPのより正確な解析結果を得ることができる。

【0106】また、上述したS比、t値、分割された集合の目的変数の平均値の差、および分割された各集合のデータ数などの値に基づいて、要因となる項目群をまたがって各説明変数の目的変数におよぼす影響の大きさ、およびその信頼度を定量的な値で評価することができる。すなわち、たとえば図20(a)に示されるようなデータのみを用いて回帰木分析をおこない、上記S比やt値や差やデータ数などの値を評価することにより、装置間差や電気的特性データというような異なる項目群の歩留りに対する影響をあらたにデータ取得することなく定量的に得ることができる。

【0107】なお、図示および詳細な説明を省略するが、トランジスタ特性を左右する重要な特性であるI-V特性についても、歩留り-電気的特性データRSPの相関関係を求める場合と同様にして、ノイズを低減させて解析をおこなうことによって、正確な解析結果を得ることができる。図21(a)はノイズの影響を除去した解析結果であり、同図(b)はノイズの影響を含む解析結果である。図20(a)および図21(a)に示すような相関図は、図1に示すデータ解析装置を用いた歩留り解析により自動的に数分以内で得ることができる。

【0108】また、図18に示す回帰木図および評価用統計値リストは、装置間差および電気的特性データの両方について歩留りに対する関係を示したものであり、装置間差よりも電気的特性データによる影響の方が小さいことを示している。そこで、装置間差の歩留りに対する影響をより正確に把握するため、たとえば図18の回帰木図中のノードn4に属するウェーハ(200枚)を除いた800枚のウェーハについてのみ回帰木分析をおこなうようにしてもよい。その場合の回帰木分析結果を図22に示す。

【0109】このようにすれば、電気的特性データ要因のうち歩留りに最も影響をおよぼす要因を取り除いて解析することができるため、装置間差がより正確に把握できるようになる。その結果、回帰木図におけるノード分岐状態が変わったり、回帰木図にはさほど変化がなくても、評価用統計値リスト中の値の有意差が小さい場合には、たとえば工程Dと工程Eのように順位が反転することも起こり得る(図18および図22参照)。なお、ノードn4に属する200枚のウェーハを除く代わりに、電気的特性データRSPが90よりも大きい、すなわちRSP>90を満たすウェーハを全て除くようにしてもよい。

【0110】図23は、オリジナルデータにおいて工程

AではAM3も使用されているとしたデータに対して回帰木分析を実行した結果の他の例を示す画面表示例である。なお、図23中、n10～n15は、説明の便宜上つけたノード番号であり、画面に表示されるわけではない。図23について説明する。1000枚のウェーハの平均歩留りは74.8%である(ノードn10)。そのうち工程AにおいてAM2またはAM3のいずれかの装置を使用した800枚のウェーハについては、その平均歩留りは80%である(ノードn12)。

【0111】ノードn12の800枚のウェーハのうち工程Cにおいて装置CM2を使用した600枚のウェーハについては、その平均歩留りは90%である(ノードn13)。ノードn13の600枚のウェーハのうち400枚のウェーハについては、電気的特性データRSPの値が90以下であり、その平均歩留りは95%である(ノードn15)。さらに、ノードn15の400枚のウェーハは、工程Aにおいて、装置AM2を使用した150枚のウェーハ(平均歩留り90%)と、装置AM3を使用した250枚のウェーハ(平均歩留り98%)に2分割される。

【0112】図23に示す例では、たとえばノードn13の600枚のウェーハを選抜し、それらについてのみ電気的特性データRSPと歩留りとの相関関係を求めたとしても、ノードn15の下にさらに工程Aの装置間差による分岐が現れているため、装置間差の影響を完全に除去することはできない。しかし、この場合でも装置間差の影響を小さくすることができるので、ノードn10の1000枚のウェーハを対象にして解析をおこなう、すなわち全く装置間差の影響を除去しないで解析するよりも、正確な解析をおこなうことができる。

【0113】上述した実施の形態によれば、回帰木分析結果として回帰木図とともに評価用統計値リストが出力されるため、その評価用統計値リストに基づいて、回帰木分析により抽出されたルールの精度や信頼度の評価をおこなうことができる。そして、評価用統計値リストは、データマイニングによる解析結果を統計的手法に不慣れたエンドユーザに納得させる際のわかりやすい判断基準といえる。

【0114】また、上述した実施の形態によれば、品種が異なっても、複数品種のデータをマージしてまとめて解析をおこなうことにより、複数品種に共通した有意差を判別することができる。それによって、たとえば、品種毎の解析では判別が困難であった各異常工程の独立性や交絡に関する情報を得ることができる。

【0115】また、上述した実施の形態によれば、異なる項目群に属する複数の項目を一括して説明変数として取り扱い、回帰木分析結果を利用して注目するノードに属する集合または属さない集合について解析をおこなうことによって、注目する項目以外の項目による影響、すなわちノイズを除去または小さくして解析をおこなうこ

とが可能となる。したがって、より正確な解析結果を得ることができ、それによってどの項目に対する改善対策が有効であるかという評価をおこなうことができる。

【0116】また、従来、判別が困難であった歩留りと電気的特性データ間の相関関係を容易に明らかにすることができる。また、この解析手法は、歩留りに効く諸々の要因の影響を低減または分離することができるので、量産段階のみならず、より複雑に諸々の要因が絡み合う開発段階においても極めて有効である。

【0117】なお、上述した実施の形態においては、半導体製造工程のプロセスデータを用いて歩留り解析をおこなう場合について説明したが、この発明は、歩留り解析に限らず、種々のデータに対する判別分析に適用することが可能である。また、上述した実施の形態においては、ルールを抽出するための統計的手法としてデータマイニング手法を用いたが、この発明はその他の有意差判定手法にも適用することができる。さらに、上述した実施の形態で用いた評価用統計値リストの各値や図8～図16、図18、図22および図23に示す回帰木図は例として示したものであり、この発明はこれらに限定されるものではない。

【0118】また、上述した実施の形態においては、電気的特性データよりも装置間差の方が歩留りに対する影響が大きい場合について説明したが、その反対に電気的特性データの方が装置間差よりも影響が大きい場合にも同様にして電気的特性データのバラツキの影響、すなわちノイズを低減した状態で装置間差の解析結果を得ることができる。この場合、回帰木図において、電気的特性データにより最上階層の分岐がなされ、装置間差による分岐はそれよりも下層で現れることとなる。

【0119】また、上述した実施の形態においては、歩留りに影響するのは装置間差と電気的特性データの2項目群であるとしたが、これに限らず、この発明は、たとえば図24に示すように、装置間差と電気的特性データとその他の3種の項目群、あるいは4種以上の項目群が歩留りに影響をおよぼす場合にも同様に適用可能である。

【0120】また、上述した実施の形態で説明したデータ解析方法は、あらかじめ用意されたプログラムをパーソナルコンピュータやワークステーション等のコンピュータで実行することにより実現することができる。また、これらのプログラムは、たとえばハードディスク、フロッピー(登録商標)ディスク、CD-ROM、MO、DVD等のコンピュータで読み取り可能な記録媒体に記録され、コンピュータによって記録媒体から読み出されることによって実行される。またこのプログラムは、上記記録媒体を介して、また、伝送媒体としてネットワークを介して配布することができる。

【0121】[付記] また、以下に示す付記1～付記21を請求項とすることができる。



【0122】（付記1）複数のオリジナルデータよりなるオリジナルデータ群と、前記オリジナルデータ間に存在するルールを抽出し、前記ルールの信頼度をあらわす情報を付加した信頼度情報付きルールファイルを作成するデータ処理部と、前記ルールファイルに基づいて前記オリジナルデータの解析をおこなう解析ツールと、を備えたことを特徴とするデータ解析装置。

【0123】（付記2）前記データ処理部は、前記オリジナルデータ群における各レコードの項目間の因果関係または影響度を明らかにし、隠された情報または規則性を含むルールを統計的手法により抽出することを特徴とする付記1に記載のデータ解析装置。

【0124】（付記3）前記データ処理部は、前記解析ツールにおいて解析すべき項目またはその組み合わせに関する指針を含むルールをデータマイニング技法により抽出することを特徴とする付記1に記載のデータ解析装置。

【0125】（付記4）前記データ処理部は、前記ルールの信頼度をあらわす情報として、複数のオリジナルデ

\*ータよりなる集合を2分割する際の分割の明確度をあらわす集合分割評価値を求めることを特徴とする付記1に記載のデータ解析装置。

【0126】（付記5）前記オリジナルデータは、変動原因または変動パターンを探ることを目的とする目的変数を含み、前記集合分割評価値は、つぎの式であらわされるS比の値であることを特徴とする付記4に記載のデータ解析装置。

$$S比 = ((S1 + S2) / 2) / S0$$

ただし、S0は、複数のオリジナルデータよりなる集合を2分割する前の目的変数の平方和、S1およびS2はそれぞれ分割後の各集合の目的変数の平方和である。

【0127】（付記6）前記オリジナルデータは、変動原因または変動パターンを探ることを目的とする目的変数を含み、前記集合分割評価値は、つぎの式であらわされるtの値であることを特徴とする付記4に記載のデータ解析装置。

【数3】

$$t = \frac{|\bar{X1} - \bar{X2}|}{\sqrt{\frac{S1 + S2}{N1 + N2 - 2} \times \left(\frac{1}{N1} + \frac{1}{N2}\right)}}$$

または、

$$t = \frac{|\bar{X1} - \bar{X2}|}{\sqrt{\frac{S1}{N1} + \frac{S2}{N2}}}$$

ただしS1およびS2は、それぞれ、複数のオリジナルデータよりなる集合を2分割してできた各集合の目的変数の平方和、N1およびN2はそれぞれ分割後の各集合の要素数、 $\bar{X1}$ および $\bar{X2}$ はそれぞれ分割後の各集合の平均である。

【0128】（付記7）前記データ処理部は、前記ルールの信頼度をあらわす情報として、2次以降の要因または条件を含む情報を求めることを特徴とする付記1に記載のデータ解析装置。

【0129】（付記8）前記データ処理部での処理対象データとして、前記オリジナルデータ群から任意のデータ種または任意のデータ項目の組み合わせを抽出する前置データ処理部をさらに備えることを特徴とする付記1～7のいずれか一つに記載のデータ解析装置。

【0130】（付記9）製造工程の歩留り低下要因を解析する装置であって、前記オリジナルデータは、変動原因または変動パターンを探ることを目的とする目的変数と、前記目的変数の変動を説明する説明変数を含み、前記目的変数は製造歩留りであり、前記説明変数は、使用装置履歴、試験結果、設計情報および測定データを含む変数であることを特徴とする付記1～8のいずれか一つに記載のデータ解析装置。

【0131】（付記10）複数のオリジナルデータ間に

存在するルールを抽出して、前記ルールにその信頼度をあらわす情報を付加した信頼度情報付きルールファイルを作成し、前記ルールとともに前記信頼度をあらわす情報を出力する工程と、前記ルールファイルに基づいて前記オリジナルデータの解析をおこなう工程と、を含むことを特徴とするデータ解析方法。

【0132】（付記11）複数の前記オリジナルデータにおける各レコードの項目間の因果関係または影響度を明らかにし、隠された情報または規則性を含むルールを統計的手法により抽出することを特徴とする付記10に記載のデータ解析方法。

【0133】（付記12）前記オリジナルデータの解析をおこなう際の解析すべき項目またはその組み合わせに関する指針を含むルールをデータマイニング技法により抽出することを特徴とする付記10に記載のデータ解析方法。

【0134】（付記13）前記ルールの信頼度をあらわす情報は、複数のオリジナルデータよりなる集合を2分

割する際の分割の明確度をあらわす集合分割評価値であることを特徴とする付記 10 に記載のデータ解析方法。

【0135】（付記 14）前記オリジナルデータは、変動原因または変動パターンを探ることを目的とする目的変数を含み、前記集合分割評価値は、つぎの式であらわされる S 比の値であることを特徴とする付記 13 に記載のデータ解析方法。

$$S \text{ 比} = ((S1 + S2) / 2) / S0$$

ただし、S0 は、複数のオリジナルデータよりなる集合\*

$$t = \frac{|\bar{X1} - \bar{X2}|}{\sqrt{\frac{S1 + S2}{N1 + N2 - 2} \times \left(\frac{1}{N1} + \frac{1}{N2}\right)}}$$

または、

$$t = \frac{|\bar{X1} - \bar{X2}|}{\sqrt{\frac{S1}{N1} + \frac{S2}{N2}}}$$

ただし S1 および S2 は、それぞれ、複数のオリジナルデータよりなる集合を 2 分割してできた各集合の目的変数の平方和、  
N1 および N2 はそれぞれ分割後の各集合の要素数、  
X1 および X2 はそれぞれ分割後の各集合の平均である。

【0137】（付記 16）前記ルールの信頼度をあらわす情報は、2 次以降の要因または条件を含むことを特徴とする付記 10 に記載のデータ解析方法。

【0138】（付記 17）ルールを抽出するための処理対象データとして、複数の前記オリジナルデータから任意のデータ種または任意のデータ項目の組み合わせをあらかじめ抽出することを特徴とする付記 10～16 のいずれか一つに記載のデータ解析方法。

【0139】（付記 18）製造工程の歩留り低下要因を解析するにあたって、前記オリジナルデータは、変動原因または変動パターンを探ることを目的とする目的変数と、前記目的変数の変動を説明する説明変数を含み、前記目的変数は製造歩留りであり、前記説明変数は、使用装置履歴、試験結果、設計情報および測定データを含む変数であることを特徴とする付記 10～17 のいずれか一つに記載のデータ解析方法。

【0140】（付記 19）前記オリジナルデータにおける各レコードは、変動原因または変動パターンを探ることを目的とする目的変数と、前記目的変数の変動を説明する説明変数として、タイプの異なる複数の項目群に属する複数の項目を有し、抽出された前記ルールに基づく回帰木図の、注目するノードに属する集合または属さない集合に対して解析をおこなうことを特徴とする付記 10～18 のいずれか一つに記載のデータ解析方法。

【0141】（付記 20）前記オリジナルデータにおける各レコードは、変動原因または変動パターンを探ることを目的とする目的変数と、前記目的変数の変動を説明する説明変数として、タイプの異なる複数の項目群に属する複数の項目を有し、抽出された前記ルールに基づく

\* を 2 分割する前の目的変数の平方和、S1 および S2 はそれぞれ分割後の各集合の目的変数の平方和である。

【0136】（付記 15）前記オリジナルデータは、変動原因または変動パターンを探ることを目的とする目的変数を含み、前記集合分割評価値は、つぎの式であらわされる t の値であることを特徴とする付記 13 に記載のデータ解析方法。

【数 4】

回帰木図の、注目するノードに属する集合または属さない集合に対して解析をおこなうことを特徴とする付記 10～18 のいずれか一つに記載のデータ解析方法。

【0142】（付記 21）前記付記 10～20 のいずれか一つに記載された方法をコンピュータに実行させるプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

【0143】

【発明の効果】この発明によれば、オリジナルデータ群の複数のオリジナルデータ間に存在するルールが抽出され、その抽出されたルールは、ルールの信頼度をあらわす情報とともに出力されるので、抽出されたルールの精度評価をおこなうことができるとともに、2 次以降の効果のある要因をその定量的な評価値とともに得ることができる。したがって、データマイニング等のデータ解析結果の精度や信頼性の評価が可能となり、産業界で広くおこなわれているデータ解析の効率化および信頼度の向上を図ることができる。この発明を製造工程の歩留り解析に適用した場合には、生産性向上に多大な効果をもたらす。また、この発明によれば、有意差の程度、複数要因の判別および説明変数間の関連性を把握することができるので、解析結果に基づいて決定された対策の効果を定量的に把握することができるという効果も得られる。

【0144】また、この発明によれば、異なる項目群に属する複数の項目を一括して説明変数として取り扱い、回帰木分析結果を利用して、注目するノードに属する集合または属さない集合について解析をおこなうことによって、ノイズがより少ない状態で解析が可能となる。したがって、注目する項目についてより正確な解析結果

を得ることができる。

【図面の簡単な説明】

【図 1】この発明にかかるデータ解析装置の構成の一例を示す概略図である。

【図 2】この発明を半導体製造工程の歩留り解析に適用した場合のデータ形式の一例を示す模式図である。

【図 3】実施の形態にかかる回帰木分析プログラムの起動時の画面表示例を示す図である。

【図 4】実施の形態にかかる回帰木分析プログラムの実行中の画面表示例を示す図である。

【図 5】実施の形態にかかる回帰木分析プログラムの実行中の画面表示例を示す図である。

【図 6】実施の形態にかかる回帰木分析プログラムの実行中の画面表示例を示す図である。

【図 7】実施の形態にかかる回帰木分析プログラムの実行中の画面表示例を示す図である。

【図 8】実施の形態にかかる回帰木分析プログラムによる分析結果を説明するための回帰木図である。

【図 9】図 8 に示す回帰木分析の結果をまとめて示す図表である。

【図 10】実施の形態において品種毎に回帰木分析を実行した結果を示す画面表示例である。

【図 11】実施の形態において品種毎に回帰木分析を実行した結果を示す画面表示例である。

【図 12】実施の形態において品種毎に回帰木分析を実行した結果を示す画面表示例である。

【図 13】実施の形態において品種毎に回帰木分析を実行した結果を示す画面表示例である。

【図 14】実施の形態において品種毎に回帰木分析を実行した結果を示す画面表示例である。

【図 15】実施の形態において多品種にわたって回帰木分析を実行した結果を示す画面表示例である。

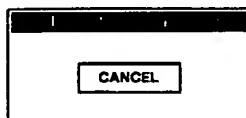
【図 16】実施の形態において多品種にわたって回帰木分析を実行した結果を示す画面表示例である。

【図 17】実施の形態においてタイプの異なる複数の項目群にまたがる回帰木分析の入力となるデータ形式の一例を示す模式図である。

【図 18】実施の形態においてタイプの異なる複数の項目群にまたがって回帰木分析を実行した結果を示す画面表示例である。

【図 7】

実施の形態にかかる回帰木プログラムの  
実行中の画面表示例を示す図



【図 19】図 18 に示す回帰木分析の結果をまとめて示す図表である。

【図 20】図 17 に示すオリジナルデータの電気的特性データ RSP と歩留りととの相関関係を装置間差の影響を低減した場合としない場合についてそれぞれ示す特性図である。

【図 21】ノイズの影響を低減した場合としない場合のトランジスタの I-V 特性の一例をそれぞれ示す特性図である。

10 【図 22】図 18 に示す回帰木分析結果に基づいて電気的特性データの影響を低減した状態で回帰木分析を実行した結果を示す画面表示例である。

【図 23】実施の形態においてタイプの異なる複数の項目群にまたがって回帰木分析を実行した結果の他の例を示す画面表示例である。

【図 24】実施の形態においてタイプの異なる複数の項目群にまたがる回帰木分析の入力となるデータ形式の他の例を示す模式図である。

20 【図 25】データマイニングを適用していない一般的なデータ解析手法を示す概念図である。

【図 26】半導体製造工程において実施されている従来のデータ解析手法を示す概念図である。

【図 27】従来の装置間差解析におけるロットの流れを示す概念図である。

【図 28】従来の装置間差解析において作成される箱ヒゲ図である。

【図 29】一般的なデータ解析において用いられるレコードの構成を示す模式図である。

30 【図 30】データマイニングを適用したデータ解析手法を示す概念図である。

【図 31】一般的な多変量解析における判別分析結果として得られるルールの精度評価について説明するための説明図である。

【符号の説明】

23 オリジナルデータ群

25 データ処理部

26 信頼度情報付きルールファイル

27 解析ツール群

29 前置データ処理部 (第 1 のデータ処理部)

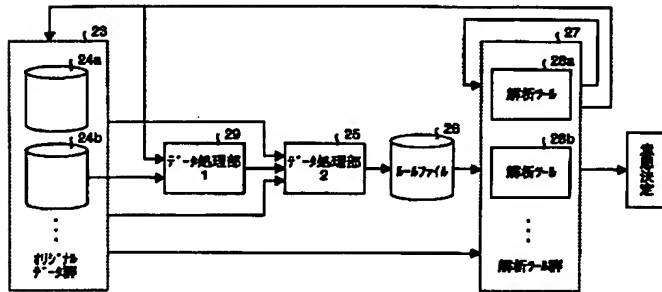
【図 17】

実施の形態においてタイプの異なる複数の項目群にまたがる  
回帰木分析の入力となるデータ形式の一例を示す模式図

ウェーハ番号	工程 A	工程 B		RSP	WETJ	歩留り
ウェーハ 1	装置 AM2	装置 BM1		評価値1	評価値1	0.806
ウェーハ 2	装置 AM1	装置 BM1		評価値2	評価値2	0.716
	装置 AM1	装置 BM2		評価値3	評価値3	0.826
ウェーハ 1	装置 AM2	装置 BM1		評価値4	評価値4	0.606

【図1】

この発明にかかるデータ解析装置の構成の一例を示す概略図



【図2】

この発明を半導体製造工程の歩留り解析に適用した場合のデータ形式の一例を示す模式図

ロット番号	工程 1	工程 2	工程 1	歩留り
ロット 1	装置 a2	装置 a3	装置 v3	0.805
ロット 2	装置 a3	装置 a5	装置 v1	0.715
	装置 a3	装置 b6	装置 v1	0.825
ロット 1	装置 a2	装置 a3	装置 v3	0.805

【図5】

実施の形態にかかる回路木分析プログラムの実行中の画面表示例を示す図

【図3】

実施の形態にかかる回路木分析プログラムの起動時の画面表示例を示す図

【図4】

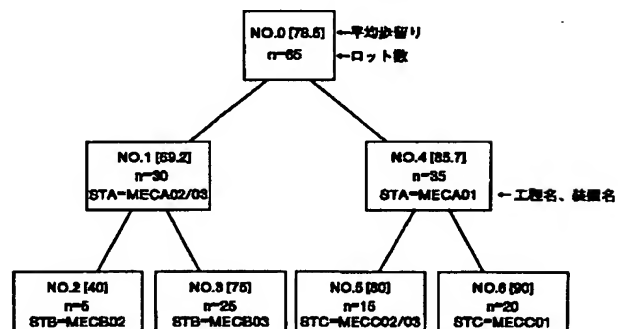
実施の形態にかかる回路木分析プログラムの実行中の画面表示例を示す図

【図6】

実施の形態にかかる回路木分析プログラムの実行中の画面表示例を示す図

【図8】

実施の形態にかかる回路木分析プログラムによる分析結果を説明するための回路木図



【図9】

図8に示す回路木 析の結果をまとめて示す図表

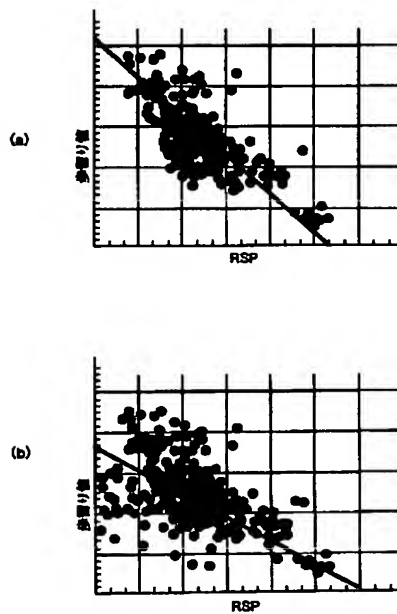
工程名			平均歩留り	ロット数
STA	STB	STC		
MECA02/03	MECB02	don't care	40%	5
	MECB03		75%	25
MECA01	don't care	MECC02/03	80%	15
		MECC01	90%	20

【図19】

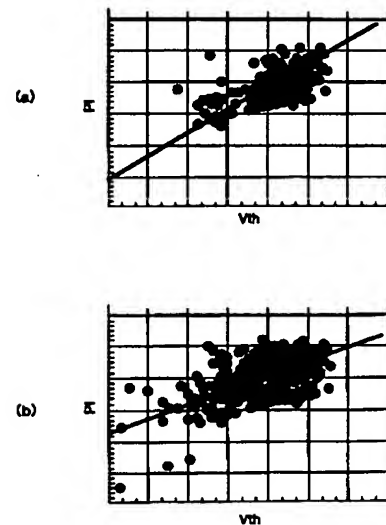
図18に示す回路木分析の結果をまとめて示す図表

工程およびウェーハ電気的特性データ				平均歩留り	ウェーハ数
工程 A	工程 B	工程 C	RSP		
AM1	BM1	don't care	don't care	45%	80
	BM3		don't care	80%	120
AM2	don't care	CM1	don't care	50%	200
		CM2	RSP>90	80%	200
			RSP≤90	95%	400

【図20】

図17に示すオリジナルデータの電気的特性データRSPと歩留りとの相関関係を  
装置間差の影響を含まない場合と含む場合についてそれぞれ示す特性図

【図21】

ノイズの影響を含まない場合と含む場合の  
トランジスタのI-V特性の一例をそれぞれ示す特性図

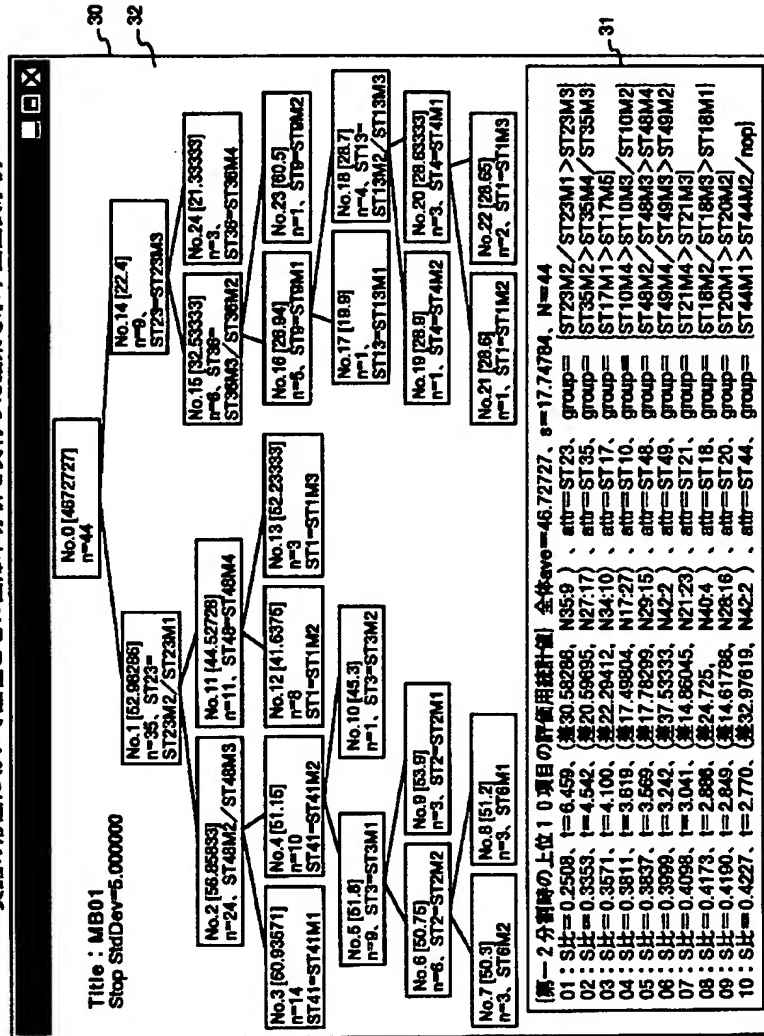
【図24】

実装の形態においてタイプの異なる複数の項目群にまたがる  
回路木分析の入力となるデータ形式の他の例を示す模式図

ウェーハ番号	工程 A	工程 B	RSP	WET_J	その他	歩留り
ウェーハ1	装置 AM2	装置 BM1	評価値1	評価値1	評価値1	0.805
ウェーハ2	装置 AM1	装置 BM1	評価値2	評価値2	評価値2	0.715
	装置 AM1	装置 BM2	評価値3	評価値3	評価値3	0.525
ウェーハ1	装置 AM2	装置 BM1	評価値4	評価値4	評価値4	0.805

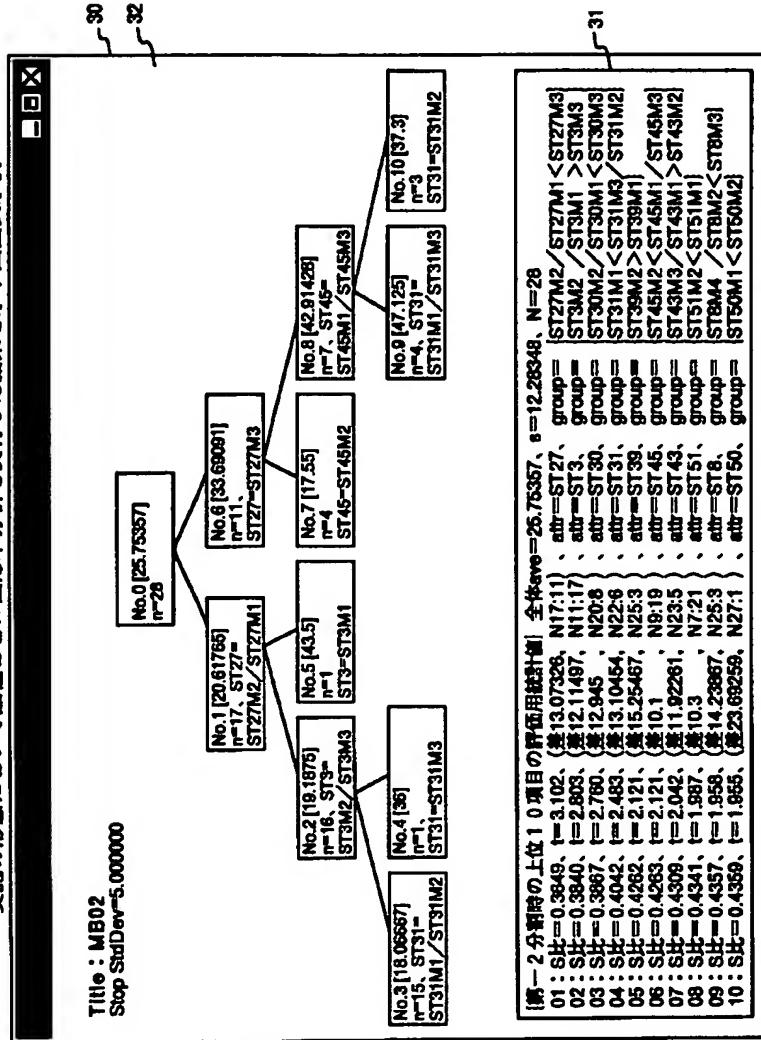
【図10】

実施の形態において品種ごとに回帰木分析を実行した結果を示す画面表示例



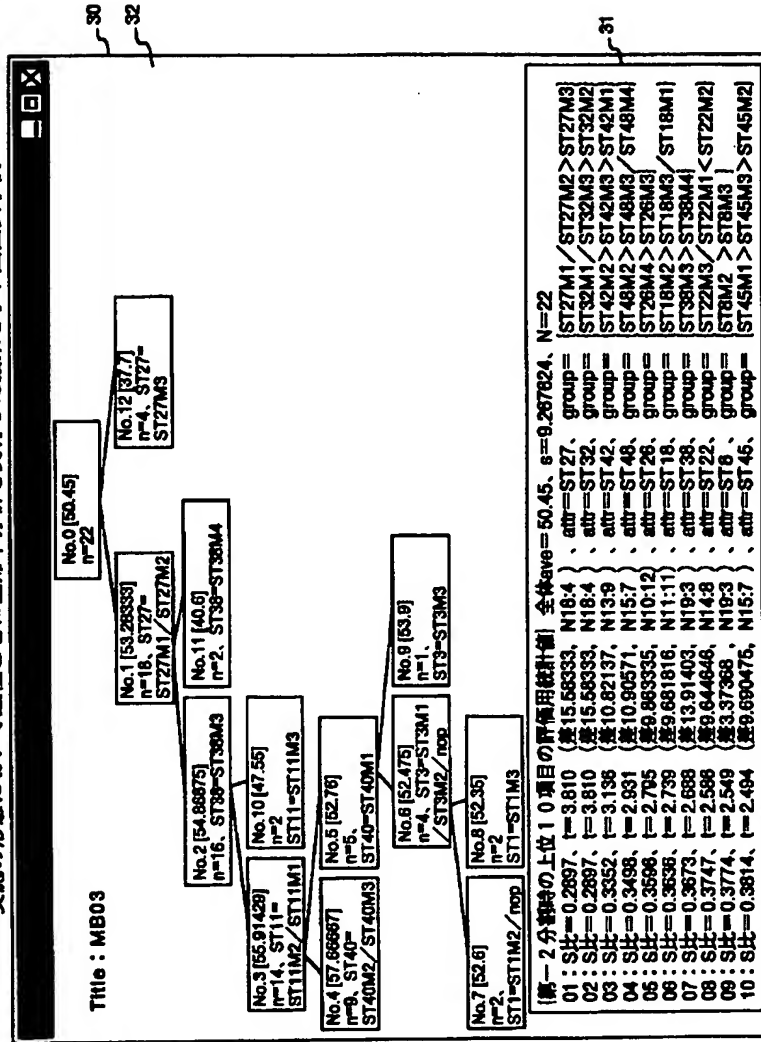
【図11】

実施の形態において品種ごとに回帰木分析を実行した結果を示す画面表示例



【図12】

実施の形態において品種ごとに四角木分析を実行した結果を示す画面表示例

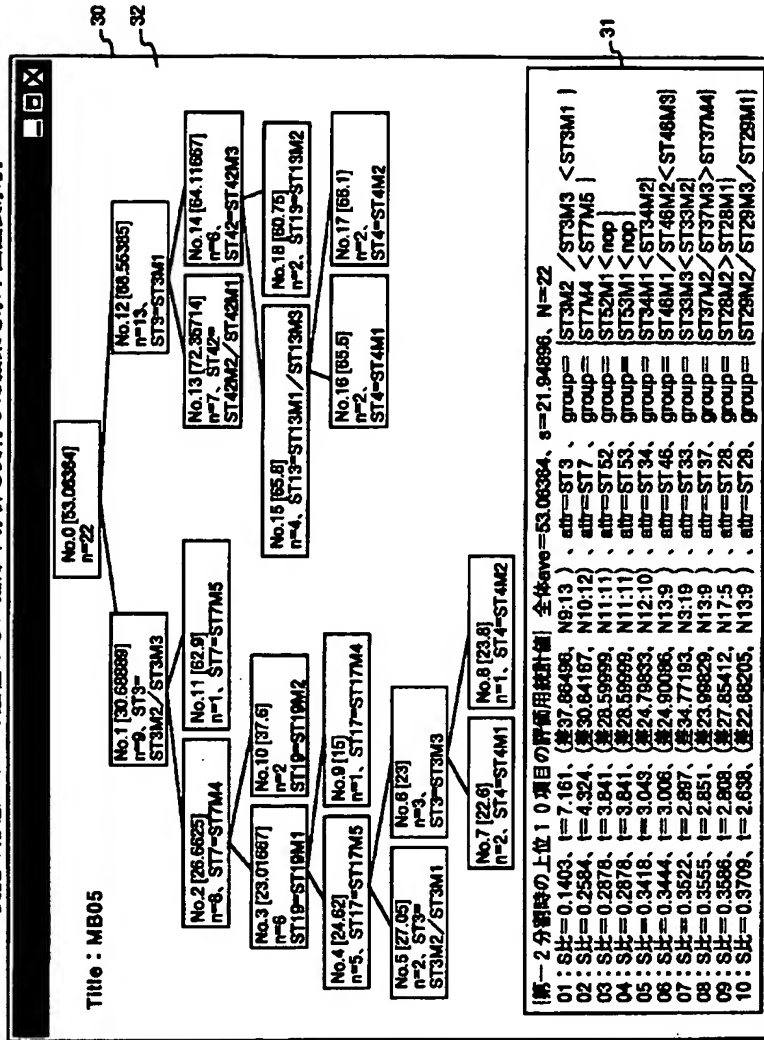




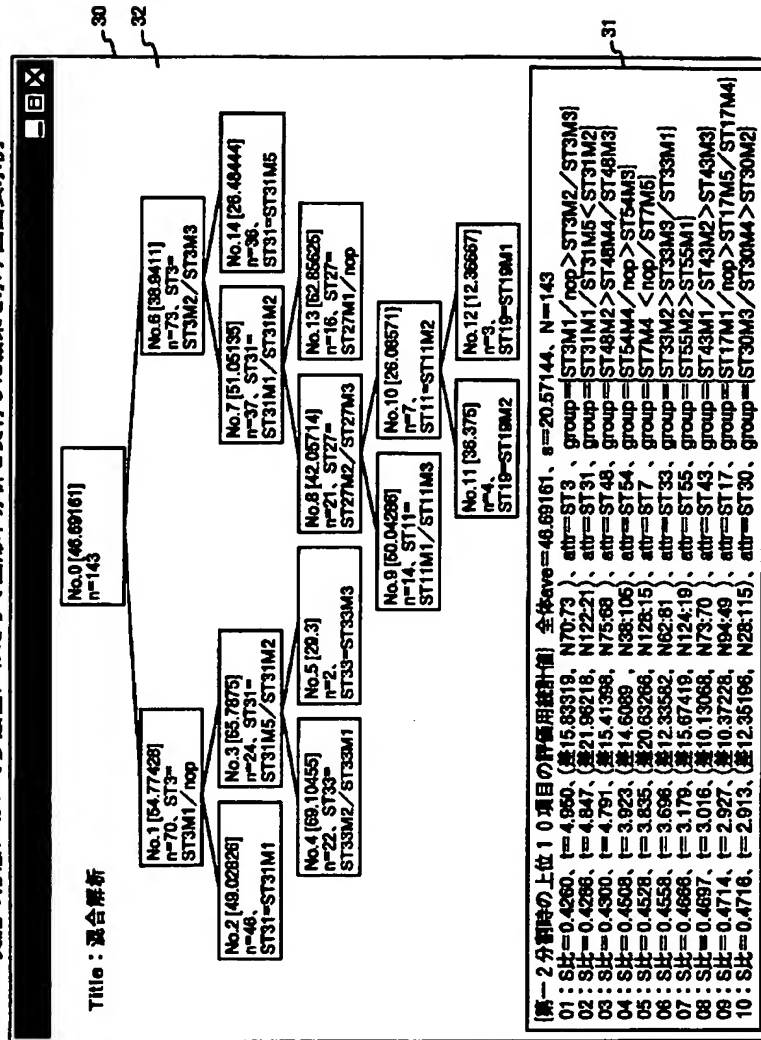


【図14】

実施の形態において品種ごとに回帰木分析を実行した結果を示す画面表示例

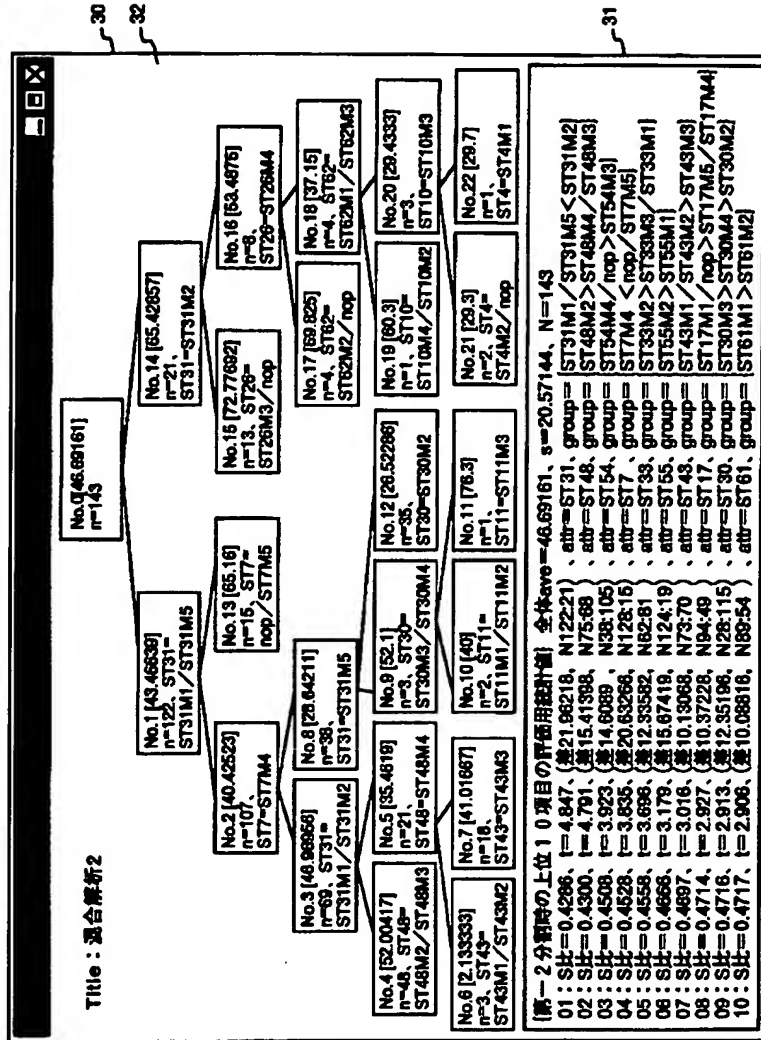


実施の形態において多品種にわたって回帰木分析を実行した結果を示す画面表示例



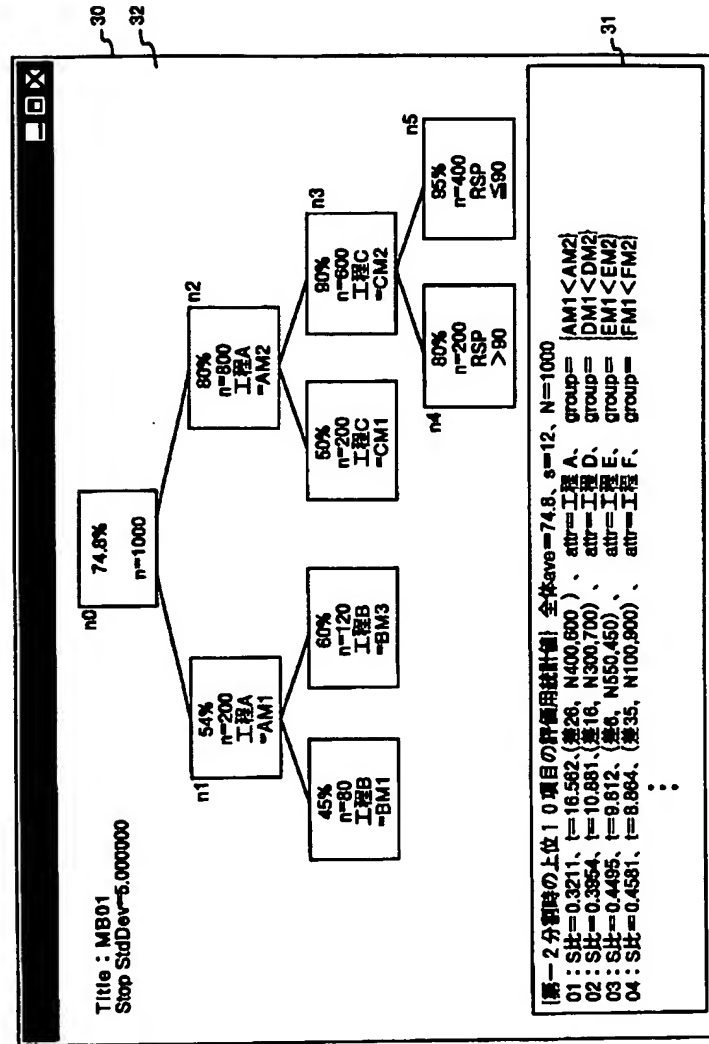
【図16】

実施の形態において多品種ごとに回帰木分析を実行した結果を示す画面表示例



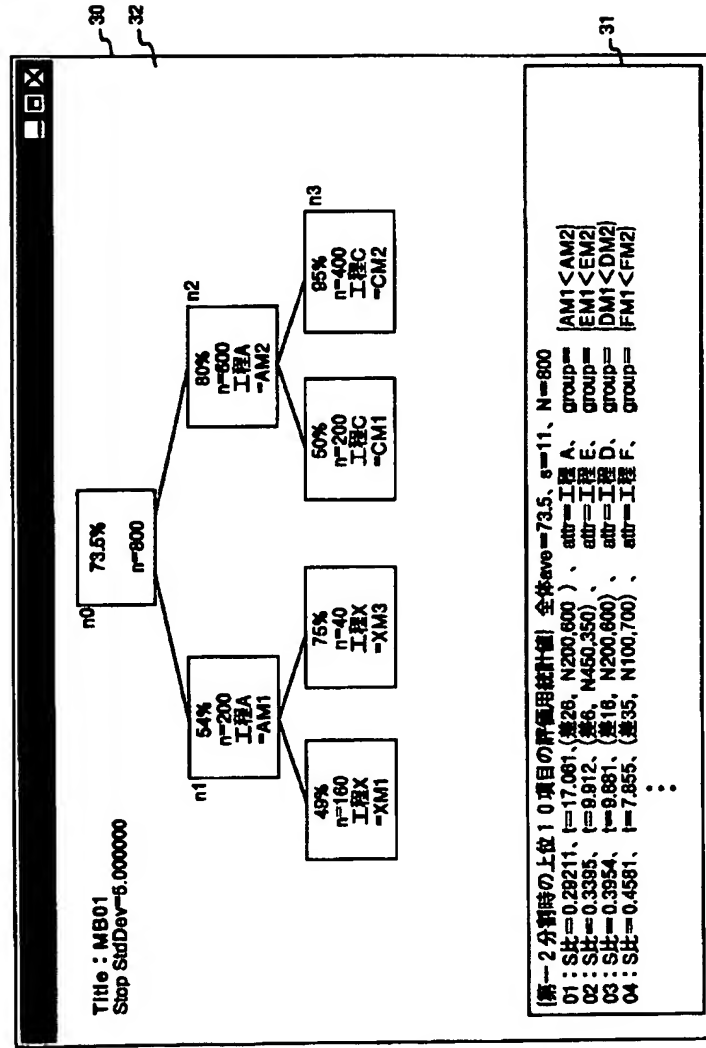
【図 18】

実施の形態においてタイプの異なる複数の項目群にまたがって  
回帰木分析を実行した結果を示す画面表示例



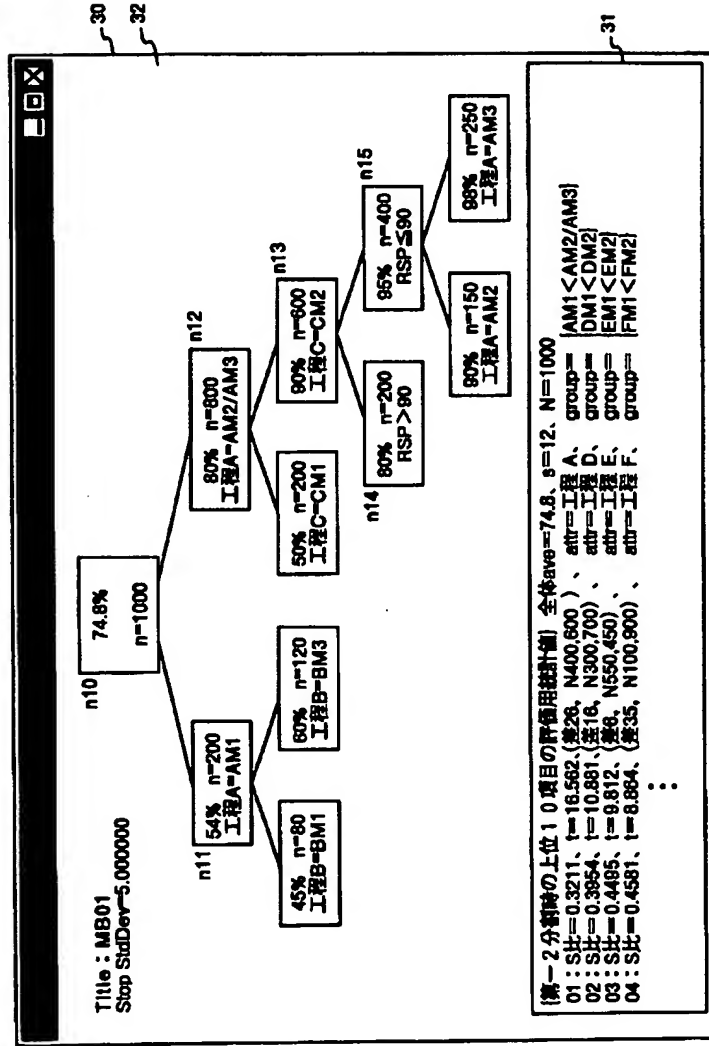
【図22】

図18に示す回帰木分析結果に基づいて電気的特性データの影響を除いた状態で回帰木分析を実行した結果を示す画面表示例



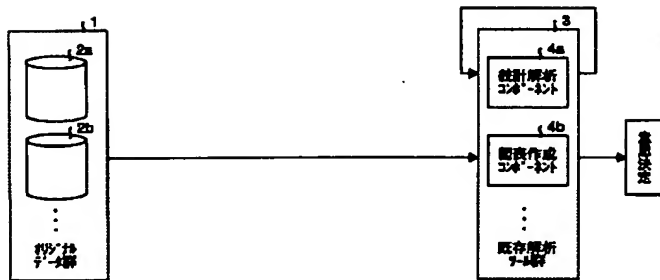
【図23】

実施の形態においてタイプの異なる複数の項目群にまたがって  
回帰木分析を実行した結果の他の例を示す画面表示例



【図25】

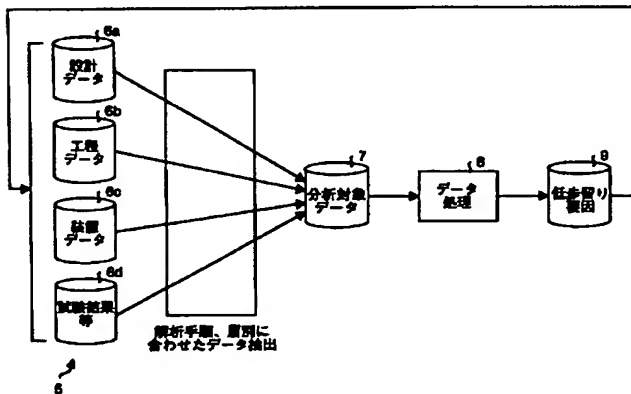
データリンクを適用していない一般的なデータ解析手法を示す概念図



【図26】

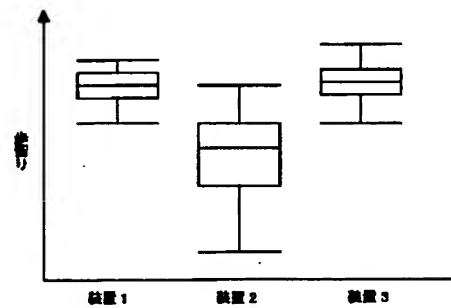
【図28】

半導体製造工程において実施されている従来のデータ解析手法を示す概念図

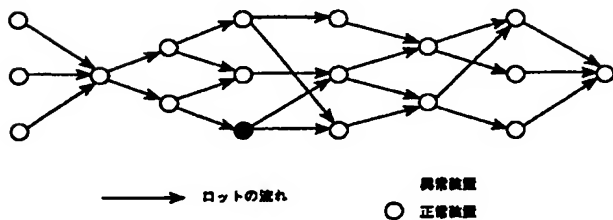


【図27】

従来の品質間差解析において作成される箱ヒゲ図



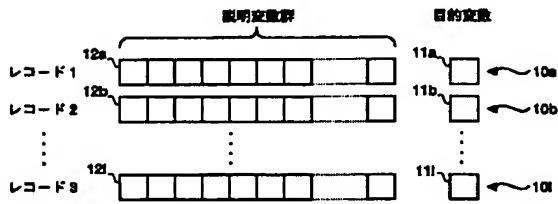
従来の品質間差解析におけるロットの流れを示す概念図





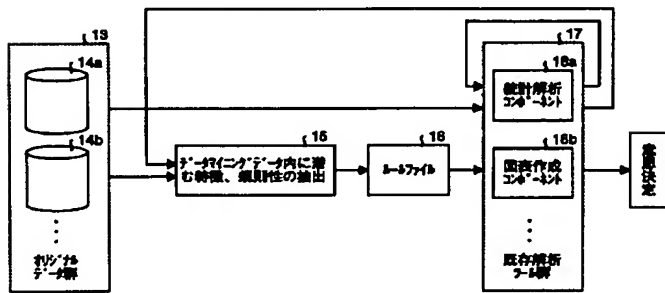
【図29】

判別分析において用いられるレコードの構成を示す模式図

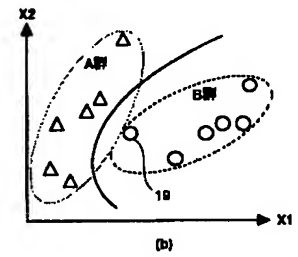
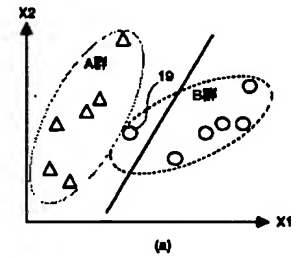


【図30】

F-統計量を適用したF-解析手法を示す概念図



【図31】

一般的な多変量解析における判別分析結果として得られる  
ルールの精度評価について説明するための説明図

フロントページの続き

(72)発明者 白井 英大  
 神奈川県川崎市高津区坂戸3丁目2番1号  
 富士通エルエスアイテクノロジー株式会社  
 内

Fターム(参考) 5B049 BB07 CC00 EE00 EE05